RUNNING HEAD: EVALUATIVE LEARNING

Expanding the Boundaries of Evaluative Learning Research: How Intersecting Regularities Shape Our Likes and Dislikes

Sean Hughes and Jan De Houwer

Ghent University

Marco Perugini

University of Milano-Bicocca

Author Note

SH and JDH, Department of Experimental Clinical and Health Psychology, Ghent University, Belgium. MP, Faculty of Psychology, University of Milan-Bicocca, Milan, Italy. Correspondence concerning this article should be sent to sean.hughes@ugent.be. We thank Maarten De Schryver for help with the statistical analyses as well as Yang Ye and Bertram Gawronski for comments on an earlier draft. The preparation of this paper was made possible by a Government of Ireland Research Fellowship to SH, Grant BOF09/01M00209 of Ghent University to JDH and VBO grant 01T02512 of Ghent University to JDH and MP.

Abstract

Over the last thirty years, researchers have identified several types of procedures via which novel preferences may be formed and existing ones altered. For instance, regularities in the presence of a single stimulus (as in the case of mere exposure) or two or more stimuli (as in the case of evaluative conditioning) have been shown to influence liking. We propose that intersections between regularities represent a previously unrecognized class of procedures for changing liking. Across four related studies, we found strong support for the hypothesis that when environmental regularities intersect with one another (i.e., share elements or have elements that share relations with other elements), the evaluative properties of the elements of those regularities can change. These changes in liking were observed across a range of stimuli and procedures and were evident when self-report measures, implicit measures (IAT) and behavioral choice measures of liking were employed. Functional and mental explanations of this phenomenon are offered followed by a discussion of how this new type of evaluative learning effect can accelerate theoretical, methodological and empirical development in attitude research.

Keywords: evaluative learning, intersecting regularities, implicit evaluations

Expanding the Boundaries of Evaluative Learning Research: How Intersecting Regularities Shape Our Likes and Dislikes

Although humans may be biologically prepared to prefer certain stimuli over others, many of our likes and dislikes are learned through on-going interactions in and with the environment (e.g., Martin & Levey, 1978). These evaluations are thought to play a causal role in a diverse spectrum of psychological phenomena, from consumer choice behaviors (Gibson, 2008; Hollands, Prestwich, & Marteau, 2011) to in-group favoritism and stigmatization (Walther, Nagengast & Trasselli, 2005), as well as self-esteem (Dijksterhuis, 2004) and voting intentions (Galdi, Arcuri & Gawronski, 2008). Preferences shape what we pay attention to (Smith, Fazio & Cejka, 1996), how we recall past events (Bohner & Dickel, 2011) as well as the judgments and decisions that we arrive at. Consequently, how, when and why likes and dislikes are established and changed is relevant for many aspects of psychological life.

Over the last thirty years, researchers have discovered several types of procedures via which novel preferences may be formed and existing ones altered (De Houwer, 2007). First, there can be regularities in the presence of a single stimulus. For instance, the mere repeated presentation of a stimulus can result in a change in liking towards that stimulus. This effect is typically referred to as the mere exposure effect (e.g., Moreland & Topolinski, 2010). Second, liking can change due to regularities in the presence of two or more stimuli (often referred to as Pavlovian contingencies). Evaluative conditioning (EC) effects, for instance, correspond to changes in liking that are due to pairing of stimuli and therefore fall into this category (for a recent review and meta-analysis see Hofmann, De Houwer, Perugini, Baeyens & Crombez, 2010). Third, regularities in the presence of a behavior and its consequences (often referred to as operant contingencies) may also generate changes in liking. For example, the physical act of pulling a lever in the presence of one stimulus and pushing a lever in the presence of another often influences how those stimuli are subsequently evaluated (i.e., approach/avoidance learning; Kawakami, Phills, Steele, & Dovidio, 2007). Finally, there is a whole range of procedures that involve the presentation of information about the attributes of stimuli, as is often the case in persuasion and impression formation research (e.g., Gawronski, Ehrenberg, Banse, Zukova, & Klauer, 2003; Smith, De Houwer, & Nosek, 2013).

Although many evaluative learning effects emerge as a function of the aforementioned procedures, several recent effects do not easily fit within this framework. To illustrate, take the self-referencing effect, which can be generated by asking participants to respond to a first neutral stimulus and self-related words (e.g., the word “me”) by pressing one response key (e.g., yellow key) and to another neutral stimulus and other-related words by pressing another response key (e.g., red key). As a result of this procedure, participants evaluate the first neutral stimulus (which was assigned to the same key as self-related words) more positively than the second neutral stimulus (see Prestwich, Perugini, Hurling, & Richetin, 2010; Perugini, Zogmaister, Richetin, Prestwich, & Hurling, 2013; see also Ebert, Steffens, von Stülpnagel, & Jelenec, 2009).

Prestwich et al. (2010) initially regarded their self-referencing effect as an instance of EC. However, upon closer inspection, stimuli in their procedure are not paired in the traditional sense of the word (i.e., co-occurrence in space and time). Instead, the observed change in liking appears to be due to an *intersection* between different environmental regularities. In the self-referencing task, the same response is correct (i.e., reinforced) when the first neutral stimulus or self-related stimuli are presented. Hence, there are two operant contingencies (i.e., "*if self-related, press the yellow key*" and "*if first neutral stimulus, press the yellow key*") that intersect in terms of a shared response and outcome (i.e., in both contingencies, pressing the yellow key is correct). Likewise, two other operant contingencies (i.e., "*if other-related, press the red key*" and "*if second neutral stimulus, press the red key*") also intersect insofar as they also share a response and outcome (i.e., in both contingencies, pressing the red key is correct). What seems to be happening here is that the valence of one stimulus (e.g., "self") transfers to another (neutral) stimulus when they both signal that the same response should be emitted (i.e., when the two stimuli are part of contingencies that intersect with regard a specific response). If this analysis has merit, then the self-referencing effect might be just one example of a much broader class of evaluative learning effects that involve changes in liking due to intersecting regularities. That is, intersecting regularities might represent a new, previously unrecognized class of procedures that offer novel ways to instigate changes in liking.

To illustrate this point more clearly, consider Figure 1, which provides an overview of all the different ways in which just one type of regularity (operant contingencies) can potentially intersect[1]. It reveals that the self-referencing effect involves only one possible intersection between operant contingencies (i.e., an intersection in terms of two elements: a common response and outcome). At the same time, it also reveals that several other intersections between operant contingencies are also possible. For instance, imagine an operant contingency in which pressing a red button (R1) in the presence of a positively valenced stimulus (S1) leads to the presentation of a neutral outcome (O1). Now imagine that

[1] The concept of intersecting regularities should not be equated with *operant contingencies*. Whereas the former refers to all possible intersections between all possible types of regularities, the latter refers to just one type of regularity. Although our argument extends to intersections within and between other types of regularities (e.g., presentations of a single stimulus or the pairing of stimuli across one or more instances of time) we have restricted our analysis in this paper solely to intersections between operant contingencies.

in a second contingency, pressing a yellow button (R2) when a neutral stimulus is present (S2) leads to the exact same outcome (O1). Participants may evaluate neutral stimulus (S2) more positively than before due to the fact that the two operant contingencies intersect in terms of a common outcome (i.e., positive stimulus (S1)→ red button (R1) → **neutral stimulus (O1)**; Neutral stimulus (S2) → yellow button (R2) → **neutral stimulus (O1)**). Figure 1 also highlights that operant contingencies can intersect not only with regard to their responses and outcomes but with regard to their antecedent stimuli as well (note that antecedent or discriminative stimuli refer to stimuli that precede the response and signal which response will likely be reinforced).

Operant Contingency 1

Operant Contingency 2	Stimulus	Response	Outcome
Stimulus	S/D	S/D	S/D
Response	S/D	S/D	S/D
Outcome	S/D	S/D	S/D

Figure 1. The potential types of intersection between stimuli, responses and outcomes that can give rise to changes in liking when two operant contingencies intersect. S stands for "same" and D stands for "different". Operant contingencies may intersect in terms of only one (e.g., their outcomes) or two features (e.g., stimuli and responses but not outcomes). The elements that intersect can either share the same function (e.g., common outcome) or have different functions in the two contingencies (e.g., when the antecedent stimulus in the first contingency is the same as the outcome in the second contingency).

A real world example might serve to illustrate this point. Imagine that you buy a new product. At home, you look for a place to store the product and notice that there is an empty shelf where you normally keep your favorite cookies. Because there is no room elsewhere, you decide to put the new product on that shelf. In abstract terms, this is one example of intersections with regard to how you behave in the context of different stimuli. More specifically, you handle a new object in a similar way (new object (S1) → put on shelf X (R1)) as you previously handled a liked object (liked object (S2) → put on shelf X (R1)). Because of the intersection between these operant contingencies in terms of a common response, a transfer in valence might occur from the liked cookies to the new object.

In short, intersections between regularities (in this case operant contingencies) might provide a previously unidentified class of procedures that can be used to modify human likes and dislikes. In the present paper, we explore this idea by examining the effects of several new procedures that all involve a particular intersection between operant contingencies. If these procedures result in changes in liking, it would not only establish each individual procedure as novel tool for changing liking but it would also support the concept of evaluative learning via intersecting regularities as a new class of evaluative learning effects. Given the sheer number of ways in which regularities can potentially intersect, we needed to restrict our initial efforts. In determining our focus, we chose the self-referencing effect as a starting point because it already seems to provide an instance of evaluative learning as the result of intersections involving one type of regularity (namely operant contingencies). Recall that the self-referencing effect involves a transfer of valence between antecedent stimuli that share a common response (e.g., self-related stimuli and those related to a first neutral concept). In Experiment 1, we examined whether a transfer of valence would occur from one stimulus to another when those stimuli were part of operant contingencies that intersected in terms of their antecedent stimuli and responses. In Experiments 2 and 3, we examined whether operant contingencies that intersect in terms of their respective outcomes would also lead to a change in liking. In Experiment 4, we looked at the impact of direct and indirect intersections between operant contingencies on changes in evaluative responding. Finally, we also examined whether evaluative learning via intersecting regularities coincides with other changes in behavior that qualify as instances of stimulus equivalence. In ways that will become apparent later, examining this question allows us to determine in a more exact manner the nature of this novel phenomenon.

Please note that the main contribution of our work is situated at a functional level of explanation, that is, the level of explanation that focusses on understanding behavior as a function of the environment. Our aim to identify a new class of procedures that can be used to change liking could indeed shed new light on those elements of the past and current environment which shape human likes and dislikes. Nevertheless, our research also has implications for the cognitive level of explanation, that is, explanations of environment-behavior relations in terms of mediating mental processes (see De Houwer, 2011, for a discussion of the relation between the functional and cognitive levels of explanation). By demonstrating the effect of various intersecting regularities on liking, we put forward a new class of evaluative learning effects that require an explanation at the cognitive level. Just like the initial demonstrations of mere exposure effects or evaluative conditioning effects stimulated fruitful debates about the mental mechanisms that mediate these effects, we hope that our research sets the stage for cognitive research on the mental processes that underlie evaluative learning via intersecting regularities. At this initial phase of research, however, we do not want to put the cart before the horse by formulating cognitive theories of a phenomenon that still needs to be established. It could well be that evaluative learning via intersecting regularities depends on the same mental processes that also drive other phenomena. In fact, at the end of our paper we will propose that known propositional learning processes might well play a crucial role in the effect of intersecting regularities on liking. But before we discuss these issues, we report four experiments that establish evaluative learning via intersecting regularities as a new type of evaluative learning effect.

Experiment 1

In our first experiment, we exposed participants to a number of fictitious brand products that had purportedly been released into the European marketplace. Participants were informed that they would encounter a number of brand names during the experiment and that they would have to taste several of these items prior to their departure. They were then administered a simple learning task. During this task they learned that pressing a button (R1) in the presence of an antecedent stimulus (S1) sometimes led to the presentation of a neutral brand name (O1) and at other times led to the presentation of a positively valenced stimulus (O2). They also learned that pressing a second button (R2) in the presence of second antecedent stimulus (S2) sometimes led to the presentation of another neutral brand name (O3) and at other times led to the presentation of a negative valenced stimulus (O4). In this way we set out to establish two different sets of operant contingencies wherein the antecedent and response of one contingency intersected with the antecedent and response of a second contingency (i.e., **verbal statement (S1)→ R1→** neutral stimulus (O1); **verbal statement (S1)→ R1 →** positive stimulus (O2) and **verbal statement (S2) → R2 →** neutral stimulus O3; **verbal statement (S2) → R2 →** negative stimulus (O4)). We predicted that these intersections between operant contingencies would give rise to a transfer of valence between the outcomes of the two contingencies. More specifically, we predicted that participants would come to like the brand name (O1) that intersected with contingencies containing positive stimuli more than the brand name (O3) that intersected with contingencies containing negative stimuli.

We probed for liking in three ways: using self-report ratings, an Implicit Association Test (IAT; Greenwald, McGhee & Schwartz, 1998), and a behavioral choice task, which was included at the end of the experiment. The IAT was added because it is thought to register a more automatic evaluation of stimuli and because it is known to be less susceptible to demand effects than self-report ratings (even though IAT effects, like virtually all measures in psychology, can be faked under certain conditions; see De Houwer, Beckers, & Moors, 2007; Röhner, Schröder-Abé & Schütz, 2013). The behavioral choice task provided a means to test whether changes in evaluative responding would also result in participants opting to approach and consume a stimulus at a later point in time.

Finally, we explored whether evaluative learning coincides with the accurate recall of previously established operant contingencies. Towards this end, participants were asked to identify what response produced a given outcome during the learning task. We focused on contingency memory because this factor has also been studied extensively and proved to be important in other types of evaluative learning (e.g., Gast, De Houwer, & De Schryver, 2012). If changes in liking due to intersecting regularities are dependent on memory for the previously established contingencies, then only those who show good memory should reveal evidence of evaluative learning on the self-report, IAT and behavioral choice tasks.

Method

Participants and design. Fifty one students at Ghent University (43 women), ranging in age from 18 to 25 years (*M* = 20.4, *SD* = 1.9) completed the study in exchange for €5 or course credit. We had to exclude the data from one participant due to technical problems during the experiment. The order of the contingency memory test (before or after the evaluative measures), the order of evaluative measures (ratings or IAT first), and the assignment of a brand name to intersecting contingencies containing positive or negative outcomes was counterbalanced across participants.

Materials

Stimuli. We used a set of six fictitious brand names as neutral stimuli (*Ailbe, Fion, Ealga, Dahy, Orin, Sile*).[2] For each participant, we selected two of these stimuli to serve as neutral stimuli during the learning phase based on how they were rated at the beginning of the experiment. A further set of sixteen positive and sixteen negative food images were also presented. All valenced images and target stimuli were presented in the center of a white screen and were 9cm wide and 8cm high. In the IAT, two fictitious brand names served as target labels and the words '*Good*' and '*Bad*' as attribute labels. Eight positively valenced and eight negatively valenced Dutch adjectives served as attribute stimuli (*delicious*, *tasty*, *nice*, *good*, *gorgeous*, *wonderful*, *yummy* and *pleasant* versus *rotten*, *disgusting*, *nasty*, *horrid*, *sick*, *vomit*, *horrible*, *unpleasant*) while images of the two brand names at different orientations served as the target stimuli.

Procedure

Upon arriving in the laboratory participants were welcomed by the researcher, asked to sign statements of informed consent and seated in front of a computer from which they received all instructions. They were informed that they would see a series of brand products that had recently been introduced into the European marketplace. They were told that it was unlikely that they had ever seen any of these brand names before, but that they would be provided with an opportunity to learn about them during the task itself. The study consisted of four main phases: stimulus pre-ratings, learning phase, a contingency memory test and evaluative measures. The entire study lasted about 30 minutes.

[2] Note that the above stimuli were in fact Gaelic (Irish) names for boys and girls. Critically, however, pre-testing indicated that none of the participants were familiar with any of these stimuli prior to the study.

Stimulus pre-ratings. Participants were informed that they would see a number of brand names and that their task was to indicate if they had encountered these names prior to the experiment. They were also asked to rate how positive or negative they considered those brand names to be using a scale ranging from -10 (Negative Feelings) to +10 (Positive Feelings) with 0 as a neutral point. For each participant, the two items closest to neutral were selected to serve as one set of outcomes during the learning phase (O1; M = 0.06 and O3; M = 1.08). Thereafter, participants rated a set of twenty delicious and disgusting looking food images using a similar scale. Based on these ratings, sixteen positively (M_{pos} = 5.87, SD_{pos} = 2.31) and sixteen negatively rated images (M_{neg} = -7.64, SD_{neg} = 1.79) were selected for inclusion in the learning phase, $t(49) = 27.43$, $p < .001$.

Stimulus		**Response**		**Outcome**
'Press J' (S1)	→	*Press J* (R1)	→	Fictitious Brand Name (O1)
'Press J' (S1)	→	*Press J* (R1)	→	Positive Stimulus (O2)
'Press F' (S2)	→	*Press F* (R2)	→	Fictitious Brand Name (O3)
'Press F' (S2)	→	*Press F* (R2)	→	Negative Stimulus (O4)

Figure 2. A visual illustration of one of the counterbalanced conditions in the learning phase in Experiment 1. Participants learned that pressing the 'J' key (R1) in the presence of the instruction 'Press J' (S1) produced either a neutral brand name (O1) or a positively valenced food image (O2) with equal probability. Likewise, they also learned that pressing the 'F' key (R2) in the presence of an instruction 'Press F' (S2) produced either a second neutral brand name (O3) or a negatively valenced image (O4) with equal probability. In this way we set out to establish two sets of operant contingencies that each intersected in terms of a common antecedent and response.

Learning phase. On-screen instructions informed participants that during the next section of the study, they would be asked to press either the 'J' or 'F' key and that this response would be immediately followed by the appearance of an image in the middle of the screen. They were asked to take their time, try to respond as accurately as possible and to remember the various images as this information would be important later on in the experiment.

The learning phase consisted of two consecutive blocks of thirty-two trials. Each trial began with the presentation of either the statement "*Please press the J key*" (S1) or "*Please press the F key*" (S2) in the middle of the screen. We chose these explicit statements in order to maximize the probability that participants would always emit the correct response. Selecting the 'J' key (R1) in the presence of (S1) resulted in the removal of the verbal statement and - following a 200ms intra-trial interval - in the presentation of a neutral brand name (O1) or one of the sixteen positively valenced food images (O2) with equal probability. Selecting the 'F' key (R2) in the presence of the second verbal statement (S2) resulted in a similar sequence of events (i.e., disappearance of the verbal statement and onset of the intra-trial interval) followed by the presentation of a second neutral brand name (O3) or a negatively valenced image (O4) with equal probability. In either case, the neutral brand name or valenced food image was displayed for 1000ms followed by a one second inter-trial interval. If participants emitted an incorrect response - such as pressing the 'F' key in the presence of (S1) or the 'J' key in the presence of (S2) – then error feedback appeared onscreen ("Wrong") and remained there until a correct response was emitted. Once a correct response was emitted the trial proceeded as normal (i.e., presentation of an outcome followed by an inter-trial interval). Therefore the learning task consisted of four different types of trials (S1 → R1 → O1; S1 → R1 → O2; S2 → R2 → O3; S2 → R2 → O4) which were presented equally often in each block. The sixteen positive and negative food images were presented once per block and the assignment of valenced images to different operant contingencies was counterbalanced across participants.

Contingency memory test. Participants were administered a short forced-choice procedure that sought to determine whether they could recall the outcome that followed a given response during the learning phase. Half of the participants were administered this test immediately after the learning phase while the other half were administered it at the end of the experiment (this way we could control for the possibility that exposure to the test altered performance on the various measures of evaluation).

Each trial presented a positive or a negatively valenced image in the middle of the screen as well as the two labels "*Press 'F' for*" and "*Press 'J' for*" on the lower left and right corners of the screen. Prior to the task, participants were informed that a brand name or food item would be displayed in the middle of the screen and that their job was to press the key that the image was paired with earlier in the study. Selecting either response removed all stimuli from the screen, followed by an inter-trial interval of 250ms and the next trial. Testing involved a block of sixteen trials, eight of which presented the two brand names (O1 or O3) while the remaining eight presented a randomly selected positive (O2) or negatively valenced image (O4). No feedback was provided for any response emitted during this task. Participants who produced a minimum of 12 out of 16 trials were defined as having passed the test while those who failed to do so were defined as having failed the task (note that if participants produced at least 12 consecutively correct responses the memory test terminated and the next phase of the experiment began). Although this pass criterion was selected arbitrarily on an *a priori* basis, we believed that responding with at least 75% accuracy would provide a useful means of distinguishing between participants who accurately recalled the contingencies and those who could not. In other words, we expected two contrasting patterns of evaluative responding based on whether participants passed or failed this test.

IAT. Prior to the onset of the IAT, participants were informed that a series of images and words would appear one-by-one in the middle of the screen and that their task was to categorize those stimuli as quickly and accurately as possible. They were also informed that the two brand names they had encountered during the learning phase (targets) as well as the words 'Good' and 'Bad' (attributes) would appear on the upper left and right sides of the screen and that stimuli could be assigned to these categories using either the left ('E') or right keys ('I'). Each trial started with the presentation of a fixation cross for 200ms in the middle of the screen followed immediately by a target or attribute stimulus. If the participant categorized the image or word correctly - by selecting the appropriate key for that block of trials - the stimulus disappeared from the screen and the next trial began. In contrast, an incorrect response resulted in the presentation of a red 'X' which remained on-screen until the correct key was pressed. Overall, each participant completed seven blocks of trials. The first block of 20 practice trials required them to sort the two brand names into their respective categories, with one brand name (O1) assigned to the left ('E') key and the other (O3) with the right ('I') key. On the second block of 20 practice trials, participants assigned positively valenced stimuli to the 'Good' category using the left key and negative stimuli to the 'Bad' category using the right key. Blocks 3 and 4 (30 trials each) involved a combined assignment of target and attribute stimuli to their respective categories. Specifically, participants categorized the first brand name (O1) and 'positive' words using the left key and the second brand name (O3) and 'negative' words using the right key. The fifth block of 20 trials reversed the key assignments, with brand name (O1) now assigned to the right key and brand name (O3) with the left key. Finally, the sixth and seventh blocks (30 trials each) required participants to categorize brand name (O1) with 'negative' words and brand name (O3) with 'positive' words.

Self-report measures. Stimulus post-ratings of the two brand names (O1 and O3) were obtained using four semantic differential scales. On each trial, participants were presented with one of the two brand names and asked to indicate whether they considered it to be '*Good/Bad*', '*Pleasant/Unpleasant*', '*Positive/Negative*' and whether '*I like it/I don't like it*' along a scale that ranged from -10 (Negative Feelings) to +10 (Positive Feelings) with 0 as a neutral point. A demand compliance check was also included to rule out the possibility that evaluative learning effects were contingent upon compliance with experimental expectations. Participants were asked to indicate the source of their evaluative responses and were given three options to choose from: (a) "*I responded to the brand names based on what I thought the experimenter wanted me to do*", (b) "*I responded to the brand names based on what I learned about them earlier in the experiment*", or (c) "*I do not know why I evaluated the brand names the way I did*". Participants were deemed demand compliant if they selected the first option and non-demand compliant if they selected from either of the later options. Nine participants were deemed demand compliant on this basis.

Behavioral choice task. Following the various measures of evaluation, the two brand names were once again printed on the computer screen. Participants were informed that samples of those brand products were waiting in an adjacent room and that their job was to indicate which of these two items they would like to taste in the final section of the experiment. After participants made their selection they were thanked, debriefed and dismissed. Note that no item was consumed during the experiment.

Results

Counterbalancing the assignment of stimuli or valence across the two contingencies as well as the order of evaluative measures and contingency memory test produced no significant effects. Consequently, analyses were collapsed across these various factors. The computer failed to record the contingency memory data of two individuals – and as such – their data was excluded from subsequent analyses. Eleven participants (23%) failed the contingency memory test while a further thirty seven (77%) passed the test.

Self-Reported Ratings

To investigate the predicted changes in liking we collapsed the four scores obtained from the stimulus post-ratings into two mean evaluative ratings – one for the first brand name (O1) and another for the second brand name (O3). We then compared these mean evaluations with those obtained for O1 and O3 prior to the learning phase (see Table 1). When these scores were submitted to a 2 (*Stimulus Type*: Brand O1 vs. O3) x 2 (*Time*: Before vs. After) x 2 (*Contingency Memory*: Pass vs. fail) repeated measures ANOVA, a significant main effect for Stimulus Type, $F(1, 46) = 7.74$, $p = .008$, $\eta^2_{partial} = .14$, a two-way interaction between Stimulus Type and Time, $F(1, 46) = 25.85$, $p < .001$, $\eta^2_{partial} = .36$, as well as a three-way interaction between Stimulus Type, Time and Contingency Memory performance was observed, $F(1, 46) = 6.15$, $p = .02$, $\eta^2_{partial} = .12$. In order to specify this three-way interaction, the impact of Stimulus Type and Contingency Memory was assessed separately for ratings obtained before and after the learning phase.

Table 1. Mean and standard deviation scores for self-reported evaluative responses as a function of contingency memory performance (pass vs. fail), test time (pre vs. post ratings) and stimulus type (brand O1 vs. O3). Note that the 'Total' column refers to scores for both pass and fail groups while * indicates that the corresponding effect differed significantly from zero ($p < .05$).

	Stimulus Pre-Ratings						Stimulus Post-Ratings					
	Pass		Fail		Total		Pass		Fail		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Self-Reported Ratings												
O1	0.24	(2.44)	-0.55	(2.54)	0.06	(2.46)	4.31*	(2.91)	1.36	(3.30)	3.64*	(3.22)
O3	1.14*	(2.23)	1.00	(3.38)	1.10*	(2.50)	-3.78*	(3.77)	-0.18	(3.56)	-2.95*	(3.99)

With respect to stimulus pre-ratings, analyses revealed a main effect for Stimulus Type, $F(1, 46) = 3.97$, $p = .05$, $\eta^2_{partial} = .08$, such that O3 was rated as slightly more positive than O1 before the learning phase (no main or interaction effects emerged for Contingency Memory performance; *ps* > .45). With respect to stimulus post-ratings, analyses revealed a main effect of Stimulus Type, $F(1, 46) = 19.88$, $p < .001$, $\eta^2_{partial} = .30$, as well as a two-way interaction between Stimulus Type and Contingency Memory performance, $F(1, 46) = 9.17$, $p = .004$, $\eta^2_{partial} = .17$. Overall, participants rated O1 more positively, and O3 more negatively following the learning phase and the difference between these post-ratings was significant, $t(47) = 6.69$, $p < .001$. Interestingly, positive evaluations of O1, $F(1, 47) = 8.17$, $p = .006$, $\eta^2_{partial} = .15$, as well as negative evaluations of O3, $F(1, 47) = 7.89$, $p = .007$, $\eta^2_{partial} = .15$, were significantly larger for those who passed the Contingency Memory test compared to those who failed. Indeed, evaluative ratings for O1 and O3 differed significantly from one another, $t(36) = 7.96$, $p < .001$, and were independently different from zero (*ps* < .001) for those in the pass group. Yet this was not the case for those who failed the Contingency Memory test, such that O1 and O3 did not occasion different ratings, $t(10) = 0.77$, $p = .46$, nor did either of those ratings differ significantly from zero (both *ps* > .2). Re-analyzing the data with demand compliant participants removed did not alter the above conclusions.

IAT

Following the recommendations of Greenwald, Nosek, and Banaji (2003), response latency data from the IAT were prepared using the D1 scoring algorithm. IAT scores reflect the difference in mean response latency between the critical blocks divided by the overall variation in those latencies. Scores were calculated so that positive values reflected a preference for the brand name that intersected with contingencies containing positive stimuli (O1) relative to the brand name that intersected with contingencies containing negative stimuli (O3). Negative values indicated the reverse pattern of responding. These scores were then submitted to a one-way ANOVA with Contingency Memory Performance as a between subjects factor.

Overall, and consistent with our expectations, participants showed an IAT effect favoring brand O1 over brand O3 (M = 0.28, SD = 0.33). Analyses also revealed a marginally significant main effect for Contingency Memory, $F(1, 47) = 2.71$, $p = .11$, $\eta^2_{partial} = .06$, such that participants who recalled the previously established contingencies showed a larger IAT effect for O1 over O3 (M = 0.32, SD = 0.34) relative to their counterparts who failed that test (M = 0.14, SD = .28). When the data of demand compliant participants were removed the obtained effects were still evident, with participants in the pass group showing a larger effect for O1 over O3 (M = 0.26, SD = .34) relative to those in the fail group (M = 0.14, SD = .28)[3].

[3] Note that IAT block order was not counterbalanced in Experiments 1 and 2 due to a programming error. As such IAT scores should not be compared to zero as a neutral reference point given that the employed block order could have produced an IAT effect that is consistent with the predicted effects of the learning task. This method factor was addressed in Experiment 3 (which was a replication of Experiment 2) as well as in Experiment 4.

Reliability Estimates

Self-reported ratings of O1 (α = .98) and O3 (α = .97), as well as the IAT scores (α = .83) appeared to have good internal consistency.

Behavioral Choice Task

Thirty three of those who passed the memory test (89%) and five of those who failed the memory test (45%) selected O1 over O3 during the choice task. When a Fisher's exact test was performed to examine whether stimulus selection was distributed differently across those who passed versus failed the contingency memory test a significant effect was obtained (p = .005). This indicates, based on the odds ratio, that the likelihood of selecting brand O1 over O3 was 9.99 times higher if participants passed the contingency memory test than if they failed to do so.

Discussion

The results of Experiment 1 reveal a transfer of valence between the outcomes of two operant contingencies that intersect with regard to their antecedent stimuli and responses. This evaluative learning effect was evident on self-reports of liking, an IAT that indexes more automatic evaluative responses and on a behavioral choice measure. Interestingly, these evaluative outcomes were for the main part evident only when participants passed a contingency memory test - regardless of whether they completed the test before or after the various measures of evaluation.

Experiment 2

Experiment 2 set out to replicate the above effects and also examine a possible boundary condition of changes in liking that occur due to intersecting regularities. In our first study, we established two operant contingencies that intersected in terms of their antecedent stimuli and responses. The self-referencing effect also involves changes in liking due to operant contingencies that intersect in two different ways (response and outcomes). Thus, until now evaluative learning effects established via intersecting regularities have only been observed when contingencies shared more than one element. But the question remains: is it possible to demonstrate a change in liking when there is only a single intersection between regularities?

With this mind, we examined whether evaluative learning can also be observed when contingencies intersect with regard to a single element (e.g., their outcomes). This time participants were informed that they would encounter a number of Chinese brand products as well as their associated ideographs. They were then exposed a learning task that sought to establish two sets of operant contingencies that intersected in terms of a common outcome. During half of the training trials participants learned that pressing a button (R1) in the presence of a positive stimulus (S1), or button (R2) in the presence of a neutral brand name (S2), caused that stimulus to disappear and a Chinese ideograph (O1) to appear onscreen. During the other half of the trials participants learned that pressing a third button (R3) in the presence of a negative stimulus (S3), or a fourth button (R4) in the presence of a neutral brand name (S4) caused that stimulus to disappear and another Chinese ideograph (O2) to take its place (see Figure 3).

Stimulus		**Response**		**Outcome**
Positive Image (S1)	→	*Response 1* (R1)	→	Neutral Ideograph (O1)
Neutral Brand (S2)	→	*Response 2* (R2)	→	Neutral Ideograph (O1)
Negative Image (S3)	→	*Response 3* (R3)	→	Neutral Ideograph (O2)
Neutral Brand (S4)	→	*Response 4* (R4)	→	Neutral Ideograph (O2)

Figure 3. A visual illustration of the learning phase in Experiment 2. Participants learned that in the presence of a positive stimulus (S1) pressing a certain button (R1) caused the positive stimulus to disappear and a Chinese ideograph (O1) to appear. Pressing a second button (R2) in the presence of a neutral brand name (S2) caused that brand to disappear and the previously mentioned ideograph (O1) to appear. Participants also learned that in the presence of a negative stimulus (S3) pressing a third button (R3) produced another Chinese ideograph (O2) while pressing a forth button (R4) in the presence of a second brand name (S4) also led to the presentation of that same ideograph O2. In this way we set out to establish two operant contingencies that intersect in terms of a common outcome.

Changes in liking were once again assessed using self-report ratings, an IAT, and a behavioral choice task. On the one hand, we predict that participants will like the neutral brand name (S2) and dislike the neutral brand name (S4) given that the former is part of a contingency that intersects with contingencies containing positive stimuli whereas the latter is part of a contingency that intersects with contingencies containing negative stimuli. On the other hand, we predict that participants will like the neutral ideograph (O1) and dislike the neutral ideograph (O2) given that the former is a member of an operant contingency containing positive stimuli while the latter is part of an operant contingency containing negative stimuli.

Stimulus Equivalence

Taking a step back, we have largely focused on a new procedure for changing likes and dislikes and have not offered an explanation for the outcomes obtained from that procedure. Drawing on over forty years' worth of research in the human learning literature, we propose that changes in liking due to intersecting regularities may be instances of a behavioral phenomenon known as stimulus equivalence (Sidman, 2009; see also Hayes, Barnes-Holmes, & Roche, 2001). Stimulus equivalence refers to the learned ability to respond to the *relation between* stimuli regardless of (a) their physical properties and in (b) ways that do not depend on direct experience or prior instruction. To illustrate, imagine that an experimenter exposes you to a procedure known as a Matching-To-Sample (MTS) task. During this task you are confronted with a 'sample' stimulus at the top of the screen (e.g., A1) and a number of 'comparison' stimuli on the bottom of the screen (e.g., B1, B2, and B3)[4]. Now imagine that you are reinforced for picking B1 whenever A1 is a sample stimulus, and on another trial, for picking C1 from a range of options (e.g., C1, C2, C3) whenever B1 is a sample stimulus (see Figure 4). Following such training people often act as if A1, B1, and C1 are related in several untrained yet predictable ways. For instance, they will tend to pick A1 rather than other available options (e.g., A2 or A3) when B1 is a sample stimulus, and this pattern of behavior is known as *symmetry*. Likewise, they will also pick A1 from a range of options when C1 is a sample stimulus, and this pattern of behavior is known as *transitivity*. Put another way, when people learn that A1 is related to B1 and B1 is related to C1 they will spontaneously act as if those three stimuli are similar or equivalent in the absence of any prior training or instruction.

[4] Note that alphanumeric symbols were used in the above example for communicative purposes. In actual equivalence research nonsense words and symbols that bear no physical similarity to one another are typically employed as sample and comparison stimuli.

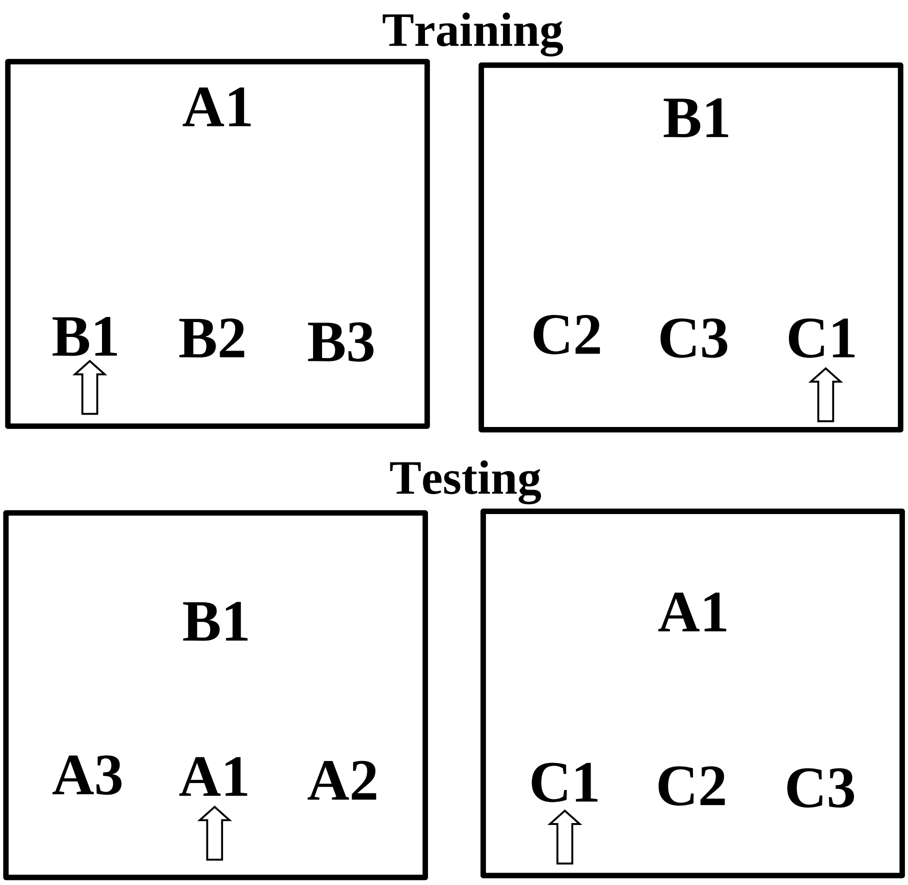

Figure 4. A visual illustration of MTS training and testing trials used to establish an equivalence relation between three arbitrary stimuli (A1, B1 and C1). Note that the arrows are included here for illustration purposes to indicate the reinforced response during training trials and the observed response during testing trials.

Stimulus equivalence is also characterized by a third important property known as the *transfer of function*. This refers to the finding that when a relationship is established between or among stimuli, the psychological properties of one stimulus may alter the properties of the other related stimuli. Imagine, for example, that A1, B1, and C1 are related as described above. Thereafter, and using a Pavlovian conditioning procedure, the experimenter repeatedly pairs A1 with an electric shock. Participants typically report fear and produce signs of physiological arousal not only for A1 but B1 and C1 as well, despite the fact that the latter stimuli were never paired with shocks in the past or share any physical similarity with A1 (e.g., Augustson & Dougher, 1997). When people respond to stimuli in these three characteristic ways (i.e., show evidence of symmetry, transitivity, and the transfer of function), those stimuli are said to participate in an equivalence relation (for a detailed treatment of equivalence and other types of relational responding see Hayes et al., 2001; Hughes & Barnes-Holmes, in press-a)[5].

Intersecting Regularities and Stimulus Equivalence

We propose that when people encounter intersections between regularities in the environment that they may treat the stimuli in those regularities as being related to one another. In the absence of contextual cues that indicate that stimuli are related in non-equivalent ways, people may come to treat the elements of intersecting regularities as equivalent to one another in some regard. Take Experiment 2. After learning the following two operant contingencies (positive stimulus (S1) → R1 → **neutral ideograph (O1)** and neutral brand name (S2) → R2 → **neutral ideograph (O1)**) participants act as if the positive stimuli, brand name and ideograph are equivalent to one another in some respect. Likewise, when they learn two additional contingencies (negative stimulus (S3) → R3 → **neutral ideograph (O2)** and neutral brand name (S4) → R4 → **neutral ideograph (O2)**) they may respond to the negative stimuli, brand name and ideograph as equivalent as well. Critically, one aspect of 'responding-as-if stimuli are equivalent' is that the evaluative properties of the valenced stimuli may transfer to the neutral stimuli with which they are related[6].

[5] Before continuing, it is important to realize that stimulus equivalence is defined functionally as a pattern of behavior rather than as a particular mental process that might explain this pattern of behavior. Hence, in line with common practice in functional psychology, when we use terms such as symmetry and transitivity, we refer to the specific ways in which people behave (i.e., the way they act as if stimuli are related in novel and untrained ways). Likewise, terms such as the transfer of function also refer to a pattern of behavior rather than mental events: people will act as if the psychological properties of those stimuli have changed in some predictable way.

[6] We do not argue that intersections between regularities will always give rise to equivalence responding (i.e., intersections are not a sufficient condition for equivalence responding). Nor is it the case that when people "act as if" stimuli are equivalent they are doing so in every conceivable sense of the word. In Experiment 2, for instance, people do not act as if brand names and valenced images look, sound, taste, or feel identical. Rather the extent to which people "act as if" stimuli are equivalent falls under the control of contextual cues that specify the

If this equivalence account of evaluative learning via intersecting regularities is correct, then changes in liking should coincide with other changes in behavior that are also part of ‘responding-as-if stimuli are equivalent’. For instance, when exposed to a MTS task with a positively valenced sample stimulus (S1) and asked to make a choice between brands (S2) and (S4), participants should select (S2) whereas they should select brand (S4) when presented with negatively valenced sample stimulus (S3). We will refer to this choice test as an ‘equivalence test’. We predict that only participants who pass this test will show an evaluative learning effect for the two brand products (S2 and S4).

Method

Participants and design. Forty one participants (34 women), ranging in age from 18 to 26 years (*M* = 20.8, *SD* = 2.0) completed the study in exchange for a monetary reward of €5 or course credit. The order of evaluative measures (self-reported vs. IAT first), as well as the assignment of neutral and valenced stimuli to different contingencies was counterbalanced across participants.

Materials

Stimuli. Participants were presented with a set of six fictitious brand names (*Ailbe, Fion, Ealga, Kapit, Orin, Sile*) and six Chinese ideographs (which were selected from those used by Payne, Cheng, Govorun, & Stewart, 2005). For each participant, we selected the two brand names and ideographs rated closest to neutral during the pre-rating phase to serve as stimuli and outcomes respectively. A further twenty positive and twenty negative food images were also pre-rated and from these ratings sixteen positively and sixteen negatively rated

---

specific ways in which stimuli are equivalent. Like all behaviors, these equivalence responses are probabilistic and the extent to which stimuli are treated as more or less equivalent will be moderated by contextual factors in the environment. The moderating impact of these contextual factors might in turn depend on a host of mediating mental processes, including the top-down regulation of attention.

images were selected for inclusion in the learning phase. Finally, two brand names or ideographs served as target labels during the IAT while the words ‘Good’ and ‘Bad’ served as attribute labels. The same positively and negatively valenced adjectives as in Experiment 1 served as attribute stimuli while images of the brand or ideographs at different orientations served as target stimuli.

Procedure

Upon arriving at the laboratory, participants were informed that they would take part in consumer research involving a series of Chinese brand names and their associated symbols. These food items were purportedly new to the marketplace - and as a result - participants were unlikely to have encountered any of these items before. Instructions also informed participants that they would have to sample some of these items at the end of the experiment. The study consisted of four main phases: stimulus pre-ratings, learning phase, an equivalence test and evaluative measures. The entire task lasted about 30 minutes.

Stimulus pre-ratings. A series of fictitious brand names and ideographs were presented onscreen one at a time and participants were asked to indicate (a) whether they had encountered those items in the past and (b) how positive or negative they considered those items to be. Thereafter, a series of valenced food images were presented on an individual basis and participants were once again asked to evaluate these items. For each participant the two brand names (S2: $M = 0.41$ and S4; $M = 0.68$) and ideographs (O1; $M = 0.17$ and O2; $M = -0.12$) with the most neutral ratings were selected as neutral stimuli while sixteen positive ($M_{pos} = 6.54$, $SD_{pos} = 1.37$) and sixteen negative images ($M_{neg} = -7.42$, $SD_{neg} = 1.48$) were selected as valenced stimuli, $t(40) = 43.32$, $p < .001$.

Learning phase. During the learning phase we sought to establish a change in liking due to operant contingencies that intersected in terms of a common outcome. On-screen instructions informed participants that they would be presented with images of food items, a Chinese symbol or a brand name. Their task was to identify the specific key (either ‘D’, ‘C’, ‘K’ or ‘M’) that a given image, brand name, or symbol was related. They were asked to take their time and try to be as accurate as possible.

Training consisted of three blocks of thirty-six trials. Each trial began with the presentation of a positively or negatively valenced stimulus (S1 or S3) or one of two brand names (S2 or S4). Selecting the ‘D’ key (R1) in the presence of a positively valenced image (S1) or the ‘C’ key when presented with a neutral brand name (S2) resulted in the removal of that stimulus from the screen, followed by a 250ms inter-stimulus interval and the subsequent presentation of a Chinese ideograph (O1). After an inter-trial interval of 1500ms the next trial began. Likewise, selecting the ‘K’ key (R3) in the presence of a negatively valenced image (S3) or the ‘M’ key when presented with a neutral brand name (S4) resulted in the removal of that stimulus from the screen, an inter-stimulus interval and the subsequent presentation of a second Chinese ideograph (O2). If participants emitted an incorrect response - such as pressing ‘M’, ‘K’ or ‘C’ in the presence of a positive image – then error feedback was displayed for 3000ms. During this time participants could not emit another response and had to wait until the next trial commenced in order to try again. Each block of trials terminated after a total of thirty-six trials or when 16 consecutively correct responses were emitted. Following each block, participants were exposed to a feedback screen that displayed their percentage accuracy during the previous section of the task. These instructions also emphasized the need for accurate responding especially if past performance was below 90%. Participants were said to have passed the learning phase if they emitted 30 out of 36 non-consecutively correct responses or 16 consecutively correct responses during the final block of the learning phase (eight participants failed to meet this criteria). Following this final block half of the participants proceeded directly to the equivalence test while the other half completed the evaluative measures.

Equivalence test. A MTS task was administered either immediately after the learning phase or at the end of the experiment in order to determine whether an equivalence relation was formed between positive stimuli (S1), ideograph (O1) and neutral brand (S2) and whether a second relation was formed between negative stimuli (S3), ideograph (O2) and neutral brand (S4) (for an overview of the MTS task, see Dymond & Whelan, 2010). Prior to the task participants were informed that "*in the next part of the experiment you should look at the item at the top of the screen, and then choose one of the items at the bottom of the screen by clicking on it with the mouse*" and that they should "*try to respond as accurately as possible*". During each test trial a "sample" stimulus (e.g., positive stimulus) was presented at the top of the screen and two "comparison" stimuli (e.g., brand names) were presented at the bottom of the screen. Selecting either of the comparison stimuli (by clicking on them with a mouse) removed all stimuli from the screen for 750ms followed by the subsequent trial. No feedback was provided for any response emitted during this task.

If the learning phase resulted in the formation of two equivalence relations as we predicted, then participants should select ideograph (O1) rather than O2 in the presence of positive images; positive rather than negative images in the presence of O1; brand (S2) rather than S4 in the presence of positive images; positive rather than negative images in the presence of S2; ideograph (O1) rather than O2 in the presence of brand (S2) and S2 rather than S4 in the presence of O1. At the same time we would also expect participants to select ideograph (O2) rather than O1 in the presence of negative images; negative rather than positive images in the presence of O2; brand (S4) rather than S2 in the presence of negative images; negative relative to positive images in the presence of S4; ideograph (O2) rather than O1 in the presence of S4 and brand (S4) rather than S2 in the presence of O2. Testing occurred across 16 trials, with each of the above trials presented in a quasi-random order. In either case, a minimum of 12 out of 16 correct responses was required to pass the test and those who did not meet this criterion were defined as having failed the test[7]**.**

IAT. Automatic evaluative responding towards stimuli that were directly (ideographs) or indirectly related (brand names) with valenced images was assessed using two IATs. In either case two of the previously encountered brand names or ideographs served as labels for the target categories as well as target category exemplars. Across both tasks, the words 'Good' and 'Bad' served as the labels of the attribute categories and eight positive and negative words as the attribute category stimuli. The presentation order of the IATs (brand versus ideograph first) was counterbalanced across participants.

Self-report measures. Stimulus-post ratings were obtained for the ideographs (O1, O2) and brand names (S2, S4) in a similar way to Experiment 1. A similar demand compliance check was also included as in Experiment 1 which identified five participants as responding in-line with perceived experimental expectations**.**

Behavioral choice task. Following the various measures of evaluation the brand names and ideographs were printed on the computer screen. Participants were informed that there were samples of these brand products in the adjacent room and that they should indicate

[7] A majority of research on stimulus equivalence repeatedly trains and tests participants until they pass the above test with 100% accuracy. One problem with such a strategy is that it becomes unclear to what extent performance on an equivalence test is driven by the untrained relations previously established during the learning phase versus the directly trained relations established by repeatedly pairing stimuli during the equivalence test (see Dymond & Rehfeldt, 2000). To protect against this possibility we exposed participants to a single round of training and testing and adopted a lower accuracy criterion on an *a priori* basis (75%). Although this pass criterion was lower than that typically seen in the literature, we believed that responding with at least 75% accuracy would provide a useful means of distinguishing between participants who act as if the stimuli are equivalent versus those who do not.

which two of these four items they would like to taste in the final section of the experiment. After participants made their selection they were thanked, debriefed and dismissed.

Results

Preliminary analyses clarified that the counterbalanced factors did not result in any main or interaction effects. Consequently, analyses were collapsed across these factors. Fourteen participants (34%) failed the equivalence test while a further twenty seven (66%) passed the test.

Self-Reported Ratings

To investigate the predicted changes in liking we calculated four mean evaluative ratings based on stimulus post-ratings - one for the ideograph directly related to positive words (O1), a second for the brand name indirectly related to positive words (S2), a third for the ideograph directly related to negative words (O2) and a forth for the brand name indirectly related with negative words (S4). We then compared these scores with the stimulus pre-ratings obtained prior to the learning phase. Submitting this data to a 4 (*Stimulus Type*: O1, O2, S2 and S4) x 2 (*Time*; before vs. after) x 2 (*Equivalence Test*: pass vs. fail) repeated measures ANOVA revealed a main effect for Stimulus Type, $F(3, 39) = 10.00$, $p < .001$, $\eta^2_{partial} = .20$, and Equivalence Test performance, $F(1, 39) = 7.45$, $p = .01$, $\eta^2_{partial} = .16$, a two-way interaction between Stimulus Type and Equivalence Test, $F(3, 39) = 25.36$, $p < .001$, $\eta^2_{partial} = .39$, Stimulus Type and Time, $F(3, 39) = 9.10$, $p < .001$, $\eta^2_{partial} = .19$, along with a three-way interaction between Stimulus Type, Time and Equivalence Test performance, $F(3, 39) = 14.30$, $p < .001$, $\eta^2_{partial} = .27$. In order to specify this three-way interaction, the impact of Stimulus Type and Equivalence Test performance was assessed separately for ratings obtained before and after the learning phase.

Table 2. Mean and standard deviation scores for self-reported evaluative responses as a function of equivalence test performance (pass vs. fail), test time (pre vs. post ratings) and stimulus type (O1, O2, S2 and S4). Note that the 'Total' column refers to scores for both pass and fail groups while * indicates that the corresponding effect differed significantly from zero ($p < .05$).

	Stimulus Pre-Ratings						Stimulus Post-Ratings					
	Pass		Fail		Total		Pass		Fail		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Self-Reported Ratings												
O1	0.00	(2.80)	0.50	(1.61)	0.17	(2.47)	5.04*	(3.93)	-0.78	(3.95)	3.05*	(4.78)
S2	0.19	(2.24)	0.86	(1.51)	0.41	(2.02)	4.82*	(3.98)	1.41	(3.38)	3.66*	(4.08)
O2	-0.67	(2.32)	0.93	(2.16)	-0.12	(2.36)	-4.06*	(4.76)	2.69*	(3.31)	-1.76*	(5.37)
S4	0.26	(2.08)	1.50*	(1.45)	0.68*	(1.97)	-5.20*	(4.07)	1.55	(3.69)	-2.89*	(5.07)

With respect to stimulus pre-ratings, analyses revealed no main or interaction effects for Stimulus Type or Equivalence Test performance (all *ps* > .06), indicating that participants did not differ in how they responded to the ideographs (O1 and O2) and brand names (S2 and S4) prior to the learning phase. With respect to stimulus post-ratings, analyses revealed a main effect of Stimulus Type, $F(3, 39) = 10.92$, $p < .001$, $\eta^2_{partial} = .22$, Equivalence Test, $F(1, 39) = 4.45$, $p = .04$, $\eta^2_{partial} = .10$, as well as a two-way interaction between Stimulus Type and Equivalence Test performance, $F(3, 39) = 22.27$, $p < .001$, $\eta^2_{partial} = .36$. Overall, participants evaluated ideograph (O1) and brand name (S2) more positively, as well as ideograph (O2) and brand name (S4) more negatively following the learning phase (see Table 2). Evaluative ratings for ideographs O1 and O2 significantly differed from one another, $t(40) = 3.25$, $p = .002$, as well as ratings for brand names S2 and S4, $t(40) = 5.04$, $p < .001$.

At the same time, and in-line with our hypothesis, these patterns of evaluative responding were closely tied to performance on the equivalence test. Participants who passed the equivalence test rated ideograph (O1) and brand name (S2) more positively, as well as ideograph (O2) and brand name (S4) more negatively, than those who failed that test (all *ps* < .01). Indeed, evaluative ratings for the ideographs (O1, O2) and brand names (S2, S4) significantly differed from one another and all evaluative scores differed from zero (*ps* < .001) when participants passed the above test. Yet this was not the case for those in the fail group. Although ratings for ideographs (O1) and (O2) differed from one another, $t(13) = 2.61$, $p = .02$, no such differences were obtained for the brand names (S2, S4) indirectly related to valenced images earlier in the experiment, $t(13) = .10$, $p = .92$. Finally, re-analyzing the data with demand compliant participants removed did not change the significance of the obtained effects.

IAT

Data from the two IATs were transformed using the D1 algorithm so that positive values reflected a response bias for the ideograph or brand name related to positive images during training. When submitted to a 2 (*IAT Type*; ideograph vs. brand names) x 2 (*Equivalence Test*) repeated measures ANOVA, a marginally significant main effect for IAT Type emerged, $F(1, 38) = 3.83$, $p = .06$, $\eta^2_{partial} = .09$, indicating that IAT effects for the brand names *indirectly* related with valenced stimuli ($M = 0.17$, $SD = 0.37$) were smaller than those obtained for ideographs *directly* related with valenced stimuli ($M = 0.31$, $SD = 0.36$). A main effect also emerged for Equivalence Test performance, $F(1, 38) = 5.89$, $p = .02$, $\eta^2_{partial} = .13$, indicating that automatic evaluative responding differed depending on whether participants treated the brand names, ideographs and valenced images as equivalent to one another. Participants in the pass group displayed a higher IAT score for ideograph (O1) relative to (O2) ($M = 0.31$, $SD = 0.39$), as well as brand name (S2) relative to (S4) ($M = 0.28$, $SD =$

0.37). Those who failed that test also showed the expected effect for ideograph (O1) over (O2) (*M* = 0.29, *SD* = 0.27), but no effect emerged for the brand names indirectly related to valenced images (*M* = -0.04, *SD* = .29). Although the two groups did not differ in their IAT scores for the ideographs, $F(1, 39) = .033$, $p = .86$, they did differ in their IAT scores for the brand names, $F(1, 40) = 7.58$, $p = .001$, $\eta^2_{partial} = .16$. Two final points are worth noting here. First, an IAT effect for ideograph (O1) over (O2) (*M* = 0.31, *SD* = .036) as well as brand (S2) over (S4) (*M* = 0.17, *SD* = 0.37) was evident even when the data from the entire sample was examined. Second, re-analyzing the data with demand compliant participants removed only served to strengthen the significance of the obtained effects, with a significant main effect emerging for both IAT Type, $F(1, 33) = 6.70$, $p = .01$, $\eta^2_{partial} = .17$, and Equivalence Test Performance, $F(1, 33) = 5.63$, $p = .02$, $\eta^2_{partial} = .15$[8].

Reliability Estimates

Self-reported ratings of O1 ($\alpha = .98$), S2 ($\alpha = .95$), O2 ($\alpha = .99$) and S4 ($\alpha = .98$), as well as the word IAT ($\alpha = .88$) and ideograph IAT scores ($\alpha = .84$) appeared to have good internal consistency.

Behavioral Choice Task

Twenty three participants who passed the equivalence test (85% of the pass group) selected the ideograph (O1) and brand name (S2) previously related to positive stimuli from the four available stimuli. In contrast, only two participants who failed the equivalence test (14% of the fail group) selected the correct stimuli while the others selected from the incorrect options. When a Fisher's exact test was performed to examine whether stimulus selection was

[8] As we mentioned previously, eight participants failed to reach the accuracy criterion during the final block of the learning phase. Removing the data of these participants and re-running the self-report and IAT analyses resulted in a similar set of findings as outlined above.

distributed differently across those who passed versus failed the equivalence test a significant effect was obtained ($p < .001$). This indicates, based on the odds ratio, that the likelihood of selecting O1 and S2 was 34.5 times higher if participants passed the equivalence test than if they failed to do so.

Discussion

Results from Experiment 2 replicate the basic findings of Experiment 1, and demonstrate that changes in liking can be engineered when operant contingencies intersect in terms of just one element. Participants learned that pressing button (R1) whenever they saw a positive stimulus (S1), or button (R2) whenever they saw neutral brand name (S2) caused that stimulus to disappear and neutral ideograph (O1) to appear. Following such training, they rated both brand name (S2) and ideograph (O1) positively. Participants also learned that pressing button (R3) whenever they saw a negative stimulus (S3) or button (R4) whenever they saw neutral brand name (S4) caused that stimulus to disappear and neutral ideograph (O2) to appear. Following such training, they rated both brand name (S4) and ideograph (O2) negatively. Comparable effects were obtained on the IAT such that an automatic bias emerged for ideograph O1 over O2 and brand product S2 over S4. This pattern of evaluative responding was also evident on the behavioral choice task with a majority of participants opting for the stimuli directly or indirectly related to positive stimuli (O1 and S2) rather those that were previous related with negative stimuli (O2 and S4). Finally, and in-line with our predictions, performance on the equivalence test was related to self-reported and automatic evaluative responding. Although participants who passed or failed the equivalence test produced comparable IAT effects for stimuli *directly* related to valenced images during the learning task, only the pass group demonstrated an IAT effect for stimuli *indirectly* related to valenced images during training. Indeed, those who failed the equivalence test reported no preferences for any stimulus, no automatic preferences for stimuli indirectly related with valenced images and selected randomly from the available options during a purported ‘taste-test’ at the end of the experiment.

Experiment 3

From the evidence presented thus far it appears that intersecting regularities represent a novel and flexible way of arranging the environment in order to modify evaluative responding. However, before we continue, it seems important to exclude an alternative and altogether more trivial explanation of the aforementioned effects. In Experiments 1-2, participants were told that they would encounter a series of novel brand products and their associated ideographs, learn about the relation between these stimuli, and later taste samples of those very items. The cover story, instructions and content of the learning tasks may have highlighted in a very salient way the intended objective of the experiment which in turn could have led participants to comply with the perceived experimental agenda (*demand compliance*) or to infer that the task set by the researcher is to identify the meaning of unfamiliar stimuli (*conversational norms*; Schwarz, 1994). We believe that the former explanation in terms of demand compliance seems unlikely because only a small subsection (nine participants in Experiment 1 and five in Experiment 2) reported that their effects were driven by demand expectations and removing their data did not influence the observed effects. Yet an explanation in terms of conversational norms does remain a possibility. Perhaps the cover story and instructions caused participants to make inferences about the task set by the researcher and thus to conform to that task. If correct, then this would imply a change in evaluative responding that is not (primarily) due to intersecting regularities but rather attempts by individuals to fulfill the role of a "*good participant*".

To exclude interpretations of our findings in terms of demand compliance or conversational norms, we decided to replicate Experiment 2 while implementing several important changes. First, we introduced a novel cover story that diverted attention away from evaluation and towards an irrelevant domain. Participants were informed that they would take part in a study examining right and left handed individuals. In order to increase the perceived validity of this cover story, the experimenter asked each participant to identify their dominant hand and then complete a short writing task that purportedly prepared them for a later section of the study. Second, in order to further conceal our intended manipulation, we informed participants that the learning task was actually a "practice task" for the main right/left handed manipulation that would take place later in the experiment. Instructions also emphasized that the assignment of neutral stimuli (words and ideographs) to different response options during this practice phase was entirely random and determined by the computer (rather than the researcher) before the experiment even began. Third, it is possible that by asking participants to pre-rate positively and negatively valenced stimuli at the beginning of the experiment, we put them in an "evaluative mindset" which influenced their subsequent behavior during the learning task. To test this possibility, we manipulated whether participants were asked to pre-rate valenced images. Finally, we included demand compliance and hypothesis awareness checks to determine whether the obtained outcomes were moderated by either variable. Observing evaluative learning effects even after such changes would provide stronger evidence that these outcomes reflect genuine changes in liking that are due to intersections between regularities.

Method

Participants and design. Eighty participants (61 women), ranging in age from 18 to 48 years (*M* = 23.1, *SD* = 5.7) completed the study in exchange for a monetary reward of €7.50. The order of evaluative measures (self-reported vs. IAT), IAT block order (consistent vs. inconsistent) and task order (ideograph vs. word), as well as assignment of neutral and valenced stimuli to different contingencies and response options was counterbalanced across participants.

Materials

Stimuli. A similar set of nonsense words, ideographs and valenced food images were employed as in Experiment 2. The two nonsense words and ideographs rated closest to neutral during the pre-rating phase were selected to serve as stimuli during the learning phase. They also served as target labels during the IAT while the words ‘Good’ and ‘Bad’ served as attribute labels. The same positively and negatively valenced adjectives as used in Experiments 1-2 served as attribute stimuli while images of the nonsense words or ideographs at different orientations served as target stimuli.

Procedure

Overall, the experiment consisted of five phases: a cover story, stimulus pre-ratings, learning phase, an equivalence test and self-report measures. The entire study lasted about 45 minutes.

Cover story. Upon arriving at the laboratory, participants were provided with a cover story which indicated that they would take part in a study investigating differences between right- and left-handed individuals. In order to increase the perceived validity of this cover story, the experimenter asked each participant to (a) state whether they were right or left handed and (b) to complete a short writing task that purportedly prepared them for later sections of the study.

Stimulus pre-ratings. For each participant, we selected the two nonsense words (S2; M = -0.14 and S4; M = 0.01) and ideographs (O1; M = 0.18 and O2; M = 0.25) that were rated closest to 0 to serve as neutral stimuli. Half of the participants were then asked to evaluate a series of valenced food images while the other half proceeded directly to the learning phase. Pre-ratings of the food images were comparable to those obtained in our earlier studies (M_{pos} = 5.80, SD_{pos} = 2.54 versus M_{neg} = -7.38, SD_{neg} = 2.06), $t(39) = 22.73$, $p < .001$.

Learning phase. Participants completed a similar learning task as in Experiment 2. However this time, onscreen instructions informed them that during the next section of the study, they would see a picture, word or symbol in the middle of the screen. Their task was to identify the specific key (either 'D', 'C', 'K' or 'M') that a given image, word, or symbol was related with. They were also informed that "*this is a practice task for the main right/left handed experiment that will take place later on. For this practice phase, the pictures, symbols, and words were chosen randomly by the computer (and assigned randomly to each of the keys) before the experiment began. So just focus on responding as accurately as possible with your left and right hands*."

Training was similar to Experiment 2 and consisted of three blocks trials that each presented an outcome when a response was emitted in the presence of a specific antecedent stimulus. In this way, we sought to establish two sets of operant contingencies that intersected in terms of a common outcome (e.g., positive stimulus (S1) → R1 → **ideograph (O1)** and nonsense word (S2) → R2 → **ideograph (O1)** vs. negative stimulus (S3) → R3→ **ideograph**

(O2) and nonsense word (S4) → R4 → **ideograph (O2)**). Each training block terminated after a total of 36 trials or when 16 consecutively correct responses were emitted. Once three training blocks had been completed participants proceeded to the equivalence test.

Equivalence test. A MTS task was administered to determine whether the expected equivalence relation was formed between positive stimuli (S1), ideograph O1 and nonsense word (S2) as well as between negative stimuli (S3), ideograph (O2) and nonsense word (S4). Testing occurred across 36 trials presented in a quasi-random order, with each of the four trained (S1→O1; S2→O1; S3→O2; S4→O2) and eight untrained stimulus relations (O1→S1; O1→S2; S1→S2; S2→S1; O2→S3; O2→S4; S3→S4 and S4→S3) examined multiple times. Individuals who scored a minimum of 30 out of 36 correct responses were defined as having passed the equivalence test while those who did not meet this criterion were defined as having failed the test.

IAT. Automatic evaluative responding towards stimuli that were directly (ideographs) or indirectly related (nonsense words) with valenced images was assessed via two separate IATs that were similar to those used in Experiment 2.

Self-report measures. Self-reported evaluative responding was assessed for the two ideographs (O1 and O2) and two nonsense words (S2 and S4) as in Experiments 1-2. We also included a similar demand compliance check as before to rule out any interpretation of the data in terms of simple compliance with experimental expectations. Ten participants were deemed demand compliant on this basis. Finally, and to exclude the possibility that our evaluative learning effects were contingent upon awareness of the experimental hypothesis, we administered a short hypothesis awareness check. Participants were asked to report what they considered the purpose of the experiment to be ("*Please indicate what you thought the* *researchers was trying to achieve in this experiment*"). They were deemed experimentally aware if they correctly identified the core parameter of the learning task (i.e., to modify liking of the neutral stimuli through their relation with valenced stimuli) and unaware of they failed to do so. Thirty one participants were deemed hypothesis aware on this basis.

Results

Preliminary analyses clarified that the counterbalanced factors did not result in any main or interaction effects. Consequently, analyses were collapsed across these various method factors. Nineteen participants (24%) failed the equivalence test while a further sixty one (76%) passed that test.

Self-Reported Ratings

Mean evaluative pre- and post-scores were calculated for the two ideographs (O1 and O2) and nonsense words (S2 and S4) as in Experiment 2. Submitting these scores to a 4 (*Stimulus Type*: O1, O2, S2 and S4) x 2 (*Time*) x 2 (*Equivalence Test*) repeated measures ANOVA revealed a main effect for Stimulus Type, $F(3, 78) = 41.34$, $p < .001$, $\eta^2_{partial} = .35$, and Equivalence Test performance, $F(1, 78) = 6.16$, $p = .02$, $\eta^2_{partial} = .07$, a two-way interaction between Stimulus Type and Equivalence Test, $F(3, 78) = 21.73$, $p < .001$, $\eta^2_{partial} = .22$, Stimulus Type and Time, $F(3, 78) = 39.79$, $p < .001$, $\eta^2_{partial} = .34$, as well as a three-way interaction between Stimulus Type, Time and Equivalence Test performance, $F(3, 78) = 20.73$, $p < .001$, $\eta^2_{partial} = .21$. We explored this three-way interaction by examining the impact of Stimulus Type and Equivalence Test performance separately for ratings obtained before and after the learning phase.

With respect to stimulus pre-ratings, analyses revealed no main or interaction effects for Stimulus Type or Equivalence Test performance (all *ps* > .19), suggesting that participants responded neutrally towards the ideographs and brand names prior to the learning phase (see Table 3). With respect to stimulus post-ratings, analyses revealed a main effect of Stimulus Type, $F(3, 78) = 46.45$, $p < .001$, $\eta^2_{partial} = .37$, and Equivalence Test performance, $F(1, 78) = 8.72$, $p = .004$, $\eta^2_{partial} = .10$, as well as a two-way interaction between Stimulus Type and Equivalence Test performance, $F(3, 78) = 24.21$, $p < .001$, $\eta^2_{partial} = .24$. Overall, participants evaluated the ideograph (O1) and nonsense word (S2) more positively, and the ideograph (O2) and nonsense word (S4) more negatively after the learning phase. Evaluative ratings for ideographs O1 and O2 significantly differed from one another, $t(79) = 9.30$, $p < .001$, as did ratings for nonsense words S2 and S4, $t(79) = 10.59$, $p < .001$.

Table 3. Mean and standard deviation scores for self-reported evaluative responses as a function of equivalence test performance (pass vs. fail), test time (pre vs. post ratings) and stimulus type (O1, O2, S2 and S4). Note that the ‘Total’ column refers to scores for both pass and fail groups while * indicates that the corresponding effect differed significantly from zero ($p < .05$).

	Stimulus Pre-Ratings						Stimulus Post-Ratings					
	Pass		Fail		Total		Pass		Fail		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Self-Reported Ratings												
O1	0.15	(2.01)	0.26	(1.66)	0.18	(1.93)	5.45*	(3.60)	1.86*	(3.89)	4.60*	(3.96)
S2	-0.34	(2.07)	0.53	(2.29)	-0.14	(2.15)	5.84*	(2.93)	1.57	(4.39)	4.83*	(3.78)
O2	0.19	(1.42)	0.42	(1.61)	0.25	(1.47)	-4.73*	(3.85)	-0.72	(4.26)	-3.78*	(4.27)
S4	-0.11	(1.65)	0.42	(2.19)	0.01	(1.79)	-5.65*	(3.26)	0.49	(4.12)	-4.19*	(4.34)

At the same time, these patterns of evaluative responding were closely related to performance on the equivalence test, such that ideograph (O1) and brand name (S2) were rated more positively, while ideograph (O2) and brand name (S4) were rated more negatively whenever participants passed the aforementioned test (all *ps* $< .001$). Indeed, evaluative ratings for the ideographs (O1 and O2), $t(60) = 10.89$, $p < .001$, as well as the nonsense words (S2 and S4), $t(60) = 15.71$, $p < .001$, significantly differed from one another and all evaluative scores differed from zero (all *ps* $< .001$) whenever participants passed the equivalence test. However, this was not the case for those who failed that same test. Neither the ideographs directly related to valenced images, $t(18) = 1.45$, $p = .17$, nor the brand names indirectly related to valenced images differed significantly from one another, $t(18) = .62$, $p = .55$, while only the ideograph (O1) directly related to positive stimuli significantly differed from zero.

IAT

Data from the two IATs were transformed and scored in a similar fashion to Experiment 2. Submitting these scores to a 2 (*IAT Type*) x 2 (*Equivalence Test*) repeated measures ANOVA revealed a main effect for Equivalence Test performance, $F(1, 78) = 10.53$, $p = .002$, $\eta^2_{partial} = .12$. Participants in the pass group showed a numerically stronger IAT effect for ideograph O1 relative to O2 ($M = 0.21$, $SD = 0.47$) compared to those who failed that test ($M = 0.01$, $SD = .43$), and in-line with Experiment 2, the difference between these two groups was not significant, $F(1, 79) = 2.79$, $p = .10$, $\eta^2_{partial} = .04$. The pass group also showed a numerically stronger IAT effect for nonsense word S2 over S4 ($M = 0.25$, $SD = 0.49$) than those in the fail group ($M = -0.13$, $SD = .38$), and consistent with Experiment 2, this difference was significant, $F(1, 79) = 9.31$, $p = .003$, $\eta^2_{partial} = .11$. Whereas the IAT effects for the ideographs and nonsense words were independently significant from zero for the pass group (both *ps* $< .001$), neither IAT effect reached significance for those that failed the equivalence test (both *ps* $> .15$). Finally, a significant effect on the ideograph ($M = 0.16$, *SD* = .46), *t*(79) = 3.12, *p* = .003, and nonsense word IAT (*M* = 0.16, *SD* = 0.49), *t*(79) = 2.83, *p* = .006, was still evident when the data from both groups was collapsed.

To conclude, participants who passed the equivalence test demonstrated clear evaluative learning effects in-line with prior training while their counterparts who failed that test showed no such effects on either self-report measures or the IAT.

Reliability Estimates

Self-reported ratings of O1 (α = .98), S2 (α = .94), O2 (α = .98) and S4 (α = .95), as well as the word IAT (α = .92) and ideograph IAT scores (α = .90) appeared to have good internal consistency.

Evaluative learning, demand compliance and hypothesis awareness

To exclude the possibility that our effects were moderated by demand compliance or awareness of the experimental hypothesis, we conducted a separate set of analyses as above with both variables as between-subject factors. No main or interaction effects for demand compliance or hypothesis awareness were observed for either the self-report or IAT data (all *ps* > .1). Indeed, re-analyzing the data with the data of the ten demand compliant participants removed did not change the significance of the obtained effects[9].

Discussion

[9] Reanalyzing the data for the 43 participants who were non-demand compliant and hypothesis unaware yielded an almost identical pattern of self-reported findings as reported before, with the main effect emerging for Stimulus Type, $F(3, 41) = 35.17$, $p < .001$, η2partial = .46, a two-way interaction between Stimulus Type and Equivalence Test, $F(3, 41) = 17.88$, $p < .001$, η2partial = .30, Stimulus Type and Time, $F(3, 41) = 30.20$, $p < .001$, η2partial = .42, as well as a three-way interaction between Stimulus Type, Time and Equivalence Test performance, $F(3, 41) = 17.94$, $p < .001$, η2partial = .30. This was also the case for the IAT data, such that a main effect was obtained for Equivalence Test performance, $F(1, 41) = 12.28$, $p < .001$, η2partial = .23, and IAT Type, $F(1, 41) = 4.68$, $p = .04$, η2partial = .10.

Experiment 3 indicates that changes in liking due to intersecting regularities are not the product of demand compliance or conversational norms established between researcher and participant. Changes in liking were still observed even when the cover story and experimental instructions directed attention away from evaluation and towards an irrelevant domain. Specifically, when two operant contingencies intersected in terms of a common outcome, the vast majority of participants treated the stimuli in those contingencies as equivalent to one another despite having received no training or instruction to do so. For instance, when a contingency containing positive stimuli intersected with a contingency containing neutral stimuli, participants self-reported and automatically evaluated stimuli in both contingencies positively (a similar effect emerged for contingencies containing negative stimuli). Finally, evaluative learning effects were not moderated by performance on either the demand compliance or experimental awareness checks.

Experiment 4

In our fourth and final study, we set out to tackle two inter-related issues. First, we sought to test another boundary condition of our account: would changes in liking due to intersecting regularities still be observed as the complexity of those intersections increase? So far we have demonstrated that simple intersections between a set of contingencies (in terms of one or two elements) is sufficient to change how previously neutral stimuli are evaluated. This time we wanted to investigate whether changes in liking would also be observed when those changes depend on both *direct* and *indirect* intersections between contingencies. To illustrate this more clearly, image that you are exposed a similar learning task as in Experiments 2 and 3. You first learn that pressing button (R1) in the presence of a neutral stimulus (S1) causes that stimulus to disappear and neutral stimulus (O1) to appear. You also learning that pressing a different button (R2) in the presence of (S1) causes that stimulus to disappear and a different neutral stimulus (O2) to appear (i.e., **S1**→ R1 → O1; **S1** → R2 → O2). In this case, you learn that two contingencies intersect *directly* in terms of a common antecedent stimulus (S1).

Thereafter you encounter a second set of intersecting contingencies. That is, when you press a certain button (R5) in the presence of a positive stimulus (S3) this stimulus disappears and a neutral stimulus (O5) takes its place. Then, when you press button (R6) in the presence of O5 that stimulus also disappears and a previously encountered neutral stimulus (O2) appears (i.e., S3 → R5 → **O5**; **O5** → R6 → O2). In this scenario, two contingencies also intersect *directly* such that the outcome of one contingency is the antecedent stimulus in another (O5). Critically, however, the first and second pair of operant contingencies also intersect *indirectly*, such that the neutral stimulus (O2) is a member of both the first and second sets of contingencies (see Figure 6).

	Stimulus	**Response**	**Outcome**
Stage 1	Neutral Stimulus (S1) →	*Response 1* (R1) →	Neutral Stimulus (O1)
	Neutral Stimulus (S1) →	*Response 2* (R2) →	Neutral Stimulus (O2)
	----------	----------	----------
	Neutral Stimulus (S2) →	*Response 3* (R3) →	Neutral Stimulus (O3)
	Neutral Stimulus (S2) →	*Response 4* (R4) →	Neutral Stimulus (O4)
Stage 2	Positive Stimulus (S3) →	*Response 5* (R5) →	Neutral Stimulus (O5)
	Neutral Stimulus (O5) →	*Response 6* (R6) →	Neutral Stimulus (O2)
	----------	----------	----------
	Negative Stimulus (S4) →	*Response 7* (R7) →	Neutral Stimulus (O6)
	Neutral Stimulus (O6) →	*Response 8* (R8) →	Neutral Stimulus (O4)

Figure 6. A visual illustration of Experiment 4's learning phase. During Stage 1 two sets of contingencies were established which intersected in terms of their respective antecedent stimuli (*S1-R1-O1 and S1-R2-O2* vs. *S2-R3-O3 and S2-R4-O4*). During Stage 2, two additional sets of contingencies were established in which the outcome of one contingency intersected with the antecedent of another (*S3-R5-O5 and O5-R6-O2* vs. *S4-R7-O6 and O6-R8-O4*).

In the above example, a change in liking towards neutral stimuli (O1) and (S1) depends on (a) a *direct* intersection between elements in one set of contingencies (i.e., two contingencies that have a common element) as well as (b) an *indirect* intersection between the first and second set of contingencies (in which the first set of contingencies intersect with the second). In other words, any observed change in positive evaluations towards S1 and O1 could only take place given that intersections within one set of contingencies share a common element with intersections that take place in another set of contingencies (note that in Experiment 4 a similar set of contingencies were also established involving negatively valenced stimuli).

A second issue we wanted to test was whether the idea of stimulus equivalence would hold for increasingly complex instances of evaluative learning. At its core, such an account argues that the above training protocol should lead to the formation of two different equivalence relations (*S1-O1-O2* and *S3-O5-O2*). Once formed, the psychological properties of positively valenced stimuli (S3) are transferred through one equivalence relation so that O2 and O5 acquire a positive valence. Given that O2 also participates in a second equivalence relation, the stimuli in this latter relation (S1 and O1) should also acquire a positive valence as well.[10]

[10] Note that it could be argued that the learning task gave rise to one rather than two separate equivalence relations given that participants come to treat those stimuli as functionally interchangeable (i.e., *S1-Positive stimuli (S3)-O1-O2-O5*).

In short, if participants learn that (a) neutral stimulus (S1) is related to neutral stimuli O1 and O2, and they then learn that (b) positive stimuli (S3) is related to neutral stimuli O2 and O5, then a transfer of valence from one equivalence relation to another should be observed given that the two relations share a common stimulus (O2). Importantly, this theoretical account places clear *a priori* restrictions on when self-reported and automatic evaluations should and should not emerge (i.e., only when participants show evidence of having formed these equivalence relations). Those who fail to do so should not show changes in self-reported or automatic evaluative responding or the tendency to approach or avoid stimuli that were indirectly related to valenced images.

Method

Participants and design. Fifty students (37 women), ranging in age from 18 to 30 years (*M* = 22.9, *SD* = 3.02) completed the study in exchange for a monetary reward or course credit. Two participants were excluded from the analysis due to errors in data recording. The order of evaluative measures, IAT block order, and assignment of neutral and positive stimuli to the various contingencies was counterbalanced across participants.

Materials

Stimuli. A set of eight nonsense words (*Jom, Sug, Cib, Vek, Hax, Teq, Zyp, Ruz*) served as target stimuli during the learning task while sixteen positive and negative food images served as positively and negative valenced stimuli respectively. Automatic evaluative responding was indexed using two variants (traditional and personalized) of the IAT. During the traditional IAT, the two stimuli (O1 and O3) that were separated from valenced images by the largest number of stimulus relations served as one set of target stimuli and the words ‘Good’ and ‘Bad’ served as another (*see below*). Participants who received the personalized

IAT were instructed to categorize stimuli based on their subjective preferences for those words. Consequently, while O1 and O3 served as target labels, the words ‘I like’ and ‘I dislike’ (rather than ‘Good’ or ‘Bad’) were employed as attribute labels. Once again, eight positive and negatively valenced adjectives served as one set of attribute stimuli while images of O1 and O3 at different orientations served as another.

Procedure

Participants completed five experimental phases: stimulus pre-ratings, learning phase, equivalence test, IATs and self-reported measures. The entire task took 45 minutes to complete.

Stimulus pre-ratings. Participants were asked to evaluate eight nonsense words (range: M = -0.69 to M = 1.33) as well as twenty positive and negative food images using a similar scale as before. From the latter ratings sixteen positive and sixteen negative stimuli were selected for inclusion in the learning phase (M_{pos} = 5.80, SD_{pos} = 1.78 versus M_{neg} = -7.08, SD_{neg} = 1.71), $t(47)$ = 29.91, p < . 001.

Learning Phase

Stage 1. During the first stage of learning, we set out to establish two sets of intersecting contingencies using a similar method as in Experiments 2-3. When participants pressed a button (R1) in the presence of a nonsense word (S1) that word disappeared and a second nonsense word (O1) appeared. Pressing a second button (R2) in the presence of the first nonsense word (S1) also caused it to disappear and a third nonsense word (O2) to appear. In this way we sought to generate two operant contingencies that intersected in terms of a common antecedent stimulus (i.e., ***S1***-*R1*-*O1* and ***S1***-*R2*-*O2*). At the same time, when participants pressed a certain button (R3) in the presence of a certain nonsense word (S2) that word disappeared and another nonsense word (O3) appeared. Pressing a forth button (R4) in the presence of nonsense word (S2) also caused it to disappear and another word (O4) to appear. In this way we sought to generate two additional operant contingencies that also intersected in terms of a common antecedent stimulus (i.e., ***S2***-*R3*-*O3* and ***S2***-*R4*-*O4*). Participants completed three blocks of thirty-five trials. A training block was terminated following (a) thirty non-consecutively correct (b) sixteen consecutively correct or (c) thirty-five trials wherein less than thirty correct responses were emitted. After the third block participants progressed to the second learning phase.

Stage 2. During the second section of the learning task two additional sets of operant contingencies were trained. When participants pressed a button (R5) in the presence of a positive stimulus (S3) that stimulus disappeared and a nonsense word (O5) appeared on screen. Pressing a second button (R6) in the presence of (O5) caused it to disappear and a previously encountered nonsense word (O2) to appear. In this way we sought to generate two operant contingencies wherein the outcome of one contingency intersected with the antecedent of the other (i.e., *S3*-*R5*-***O5*** and ***O5***-*R6*-*O2*). At the same time, when participants pressed a button (R7) in the presence of a negative stimulus (S4) that stimulus disappeared and a nonsense word (O6) appeared. Pressing another button (R8) in the presence of (O6) caused it to disappear and a previously encountered word (O4) to appear. In this way we sought to generate two operant contingencies wherein the outcome of one contingency intersected with the antecedent of the other (i.e., *S4*-*R7*-***O6*** and ***O6***-*R8*-*O4*).

Stage 3. In the final learning phase we exposed participants to a single block of thirty-six trials from both of the previous stages. Drawing on what they had learned during stages one and two, participants were required to indicate which of the eight available response options a given stimulus was paired with.

Equivalence test. We administered a series of equivalence tests during the learning phase in order to examine whether the various relations were formed as predicted. More precisely, we assessed the formation of two equivalence relations (*S1-O1-O2* and *S2-O3-O4*) at the end of Stage 1, the formation of another two relations (*Positive stimuli (S3)-O5-O2* and *Negative stimuli (S4)-O6-O4*) at the end of Stage 2 and the combination of those relations at the end of Stage 3. Participants who responded with a minimum of 12 out of 16 correct responses in the final equivalence test were defined as having passed the test while those who did not meet this criterion were defined as having failed. No feedback was provided for any response emitted during this task.

IAT. To demonstrate the generalizability of our evaluative learning effects, we asked each participant to complete two (traditional and personalized) variants of the IAT. For the traditional IAT, the two stimuli furthest removed (relationally) from valenced images during training (O1 and O3) served as one set of target labels while the Dutch words for ‘Good’ and ‘Bad’ served as another. The personalized IAT was identical to the traditional IAT with three exceptions. First, participants were asked to categorize O1 and O3 based on their subjective impressions of those stimuli; thus the target labels ‘I like it’ and ‘I dislike it’ were employed in place of the generic words ‘Good’ and ‘Bad’. Second, and in-line with Olson and Fazio (2004), no error feedback was provided given that participants were responding based on their personal evaluations of the presented stimuli. Third, and to minimize fatigue effects, the personalized IAT was shortened so that practice blocks consisted of sixteen trials while test blocks consisted of thirty-two trials. Within each block the target and attribute stimuli were presented randomly in alternating order while the presentation and critical block order of the two IATs (consistent versus inconsistent first) was counterbalanced across participants and remained the same during both versions of the task.

Self-report measures. Self-reported evaluative responding was assessed for the eight previously neutral stimuli (S1, S2, O1, O2, O3, O4, O5 and O6) as in Experiments 1-2. A demand compliance check revealed that five participants responded in-line with the perceived experimental agenda.

Behavioral choice task. Following the self-report and IAT measures the eight previously neutral stimuli were printed on the computer screen. Participants were informed that in the adjacent room there were samples of the various brand products that they had encountered earlier in the experiment. Their job was to indicate which four of these items they would like to taste in the final section of the experiment. After participants made their selection they were thanked, debriefed and dismissed.

Results

Preliminary analyses indicated that only block order influenced performance on the IAT. Consequently, analyses were collapsed across all other method factors. Sixteen participants (33%) failed the final equivalence test while a further thirty two (67%) passed the test.

Self-Reported Ratings

A series of evaluative pre- and post-scores were calculated for each of the previously neutral stimuli (S1, S2, O1, O2, O3, O4, O5 and O6) in a similar manner to Experiments 1-3. Submitting these scores to a 8 (*Stimulus*) x 2 (*Time*) x 2 (*Equivalence Test*) repeated measures

ANOVA revealed a main effect for Stimulus Type, $F(7, 46) = 14.51$, $p < .001$, $\eta^2_{partial} = .24$, as well as a two-way interaction between Stimulus Type and Equivalence Test, $F(7, 46) = 5.79$, $p < .001$, $\eta^2_{partial} = .11$, Stimulus Type and Time, $F(7, 46) = 10.86$, $p < .001$, $\eta^2_{partial} = .19$, and a three-way interaction between Stimulus Type, Time and Equivalence Test performance, $F(7, 46) = 5.69$, $p < .001$, $\eta^2_{partial} = .11$. We examined the three-way interaction by assessing the impact of Stimulus Type and Equivalence Test performance separately for ratings obtained before and after the learning phase.

With respect to stimulus pre-ratings, analyses revealed a main effect for Stimulus Type, $F(7, 46) = 3.02$, $p = .004$, $\eta^2_{partial} = .06$, and Equivalence Test performance, $F(1, 46) = 7.29$, $p = .01$, $\eta^2_{partial} = .14$. Participants who would subsequently pass the equivalence test evaluated stimuli slightly more positively than those who would later fail that test (see Table 4). Critically, however, the evaluative ratings for pass and fail groups did not differed from one another (all *ps* > .60) with a single exception (O3; $F(1, 47) = 6.46$, $p = .01$, $\eta^2_{partial} = .12$). Generally speaking, then, it appears that participants rated the various stimuli neutrally prior to the learning phase.

With respect to stimulus post-ratings, analyses revealed a main effect of Stimulus Type, $F(7, 46) = 15.88$, $p < .001$, $\eta^2_{partial} = .26$, as well as a two-way interaction between Stimulus Type and Equivalence Test performance, $F(7, 46) = 7.08$, $p < .001$, $\eta^2_{partial} = .13$. Overall, participants evaluated stimuli that had been directly (O5) and indirectly (S1, O1, O2) related to positive images as positive while stimuli that had been directly (O6) or indirectly (S2, O3, O4) related to negative images were evaluated negatively. They also demonstrated the expected pattern of preferences for stimulus O5 over O6, $t(47) = 6.86$, $p < .001$, S1 over S2, $t(47) = 4.27$, $p < .001$, O1 over O3, $t(47) = 4.19$, $p < .001$, and O2 over O4, $t(47) = 5.07$, *p*

< .001, and these ratings were all independently different from zero (*ps* < .01) with the exception of S2, $t(47) = 1.37$, $p = .18$).

Table 4. Mean and standard deviation scores for self-reported evaluative responses as a function of equivalence test performance (pass vs. fail), test time (pre vs. post ratings) and stimulus type (S1, S2, O1, O2, O3, O4, O5 and O6). Note that the 'Total' column refers to scores for both pass and fail groups while * indicates that the corresponding effect differed significantly from zero ($p < .05$).

	Stimulus Pre-Ratings						Stimulus Post-Ratings					
	Pass		Fail		Total		Pass		Fail		Total	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Self-Reported Ratings												
S1	1.34*	(2.77)	1.31	(2.62)	1.33*	(2.70)	4.56*	(4.49)	3.06*	(4.11)	4.06*	(4.38)
O1	0.69	(2.76)	-0.56	(2.65)	0.27	(2.77)	4.63*	(3.82)	0.50	(4.60)	3.25*	(4.49)
O2	0.16	(1.96)	-0.81	(2.63)	-0.17	(2.23)	3.37*	(4.48)	0.19	(4.58)	2.31*	(4.71)
O5	0.31	(2.98)	-1.31	(2.85)	-0.23	(3.01)	4.94*	(3.83)	2.13	(4.88)	4.00*	(4.37)
S2	-0.09	(2.16)	0.75	(2.89)	0.19	(2.43)	-2.87*	(5.32)	2.38	(4.79)	-1.13	(5.68)
O3	0.56	(2.68)	-1.44*	(2.34)	-0.10	(2.72)	-3.00*	(5.39)	-0.44	(5.62)	-2.15*	(5.55)
O4	0.81	(2.90)	-0.63	(2.58)	0.33	(2.85)	-3.78*	(3.37)	-0.19	(4.09)	-2.58*	(3.97)
O6	-0.25	(3.04)	-1.56*	(2.45)	-0.69	(2.90)	-5.78*	(3.94)	-1.69	(5.77)	-4.42*	(4.97)

Similar to Experiments 2-3, we found that the above changes in liking were closely tied to equivalence test performance. Participants in the pass group tended to evaluate items related to positive stimuli more positively as well as items related to negative stimuli more negatively compared to their counterparts who failed that same test (all *ps* < .03) (note that there were two exceptions; S1; $F(1, 47) = 1.26$, $p = .27$, and O3; $F(1, 47) = 2.34$, $p = .13$). The pass group also showed the expected pattern of preferences for O5 over O6, $t(31) = 8.77$, $p < .001$, S1 over S2, $t(31) = 5.01$, $p < .001$, O1 over O3, $t(31) = 5.19$, $p < .001$, and O2 over O4, $t(29) = 6.31$, $p < .001$, while all ratings independently differed from zero (*ps* < .005). In contrast, the fail group showed none of the expected patterns of relative preferences (all *ps* > .14) while the only the rating for S1 significantly differed from zero, $t(15) = 2.98$, $p = .009$. Finally, re-analyzing the data with the five demand compliant participants removed did not influence the above effects.

IAT

Data from the two IATs were transformed as in Experiments 1-3 and scored so that positive values reflected a bias for the stimuli furthest removed from positive images during training (O1 relative to O3). When submitted to a 2 (*IAT Type*; Traditional vs. Personalized) x 2 (*Equivalence Test*) x 2 (*Block Order*) repeated measures ANOVA, a main effect emerged for Block Order, $F(1, 42) = 4.16$, $p = .05$, $\eta^2_{partial} = .09$, and Equivalence Test performance, $F(1, 42) = 5.89$, $p = .02$, $\eta^2_{partial} = .12$, along with a marginally significant three-way interaction between IAT Type, Block Order and Equivalence Test performance, $F(1, 42) = 3.78$, $p = .06$, $\eta^2_{partial} = .08$. In order to facilitate interpretation of this three-way interaction, we examined the impact of Block Order and Equivalence Test performance separately for scores obtained from the personalized and traditional IATs.

With respect to the traditional IAT, analyses revealed a main effect of Equivalence Test performance, $F(1, 45) = 5.66$, $p = .02$, $\eta^2_{partial} = .12$, and a marginally significant two-way interaction between Equivalence Test and Block Order, $F(1, 45) = 3.54$, $p = .07$, $\eta^2_{partial} = .08$. Consistent with our predictions, participants who acted as if stimuli were equivalent showed an automatic evaluative bias for O1 relative to O3, regardless of whether they completed the IAT in a consistent ($M = 0.10$, $SD = .40$) or inconsistent first order ($M = 0.23$, $SD = 0.50$). Participants who failed the equivalence test showed no evidence of an IAT effect when they encountered a consistent first block order ($M = 0.02$, $SD = 0.38$) and an opposite evaluative bias for O3 over O1 when they encountered an inconsistent first block order (*M* = -0.41, *SD* = 0.57). Although the traditional IAT score for the pass group differed significantly from zero, *t*(30) = 2.19, *p* = .04, no such effect were observed in the fail group, *t*(14) = 0.97, *p* = .35. With respect to the personalized IAT only a main effect for Block Order was observed, *F*(1, 47) = 5.20, *p* = .03, $\eta^2_{partial}$ = .11, with participants showing a slightly larger effect following inconsistent (*M* = -0.16, *SD* = 0.40) compared to a consistent first block order (*M* = 0.09, *SD* = 0.32). Critically, however, neither effect was significantly different from zero, (both *ps* > .30). Re-analyzing the data with the five demand compliant participants removed did not alter these general conclusions.

In short, evidence for automatic evaluative responding was obtained on the traditional but not the personalized IAT, and these performances varied as a function of IAT block order as well as the extent to which participants treated stimuli as equivalent to one another.

Reliability Estimates

The traditional IAT (α = .93) and personalized IAT scores (α = .81) appeared to have good internal consistency.

Behavioral Choice Task

A correct response on the behavioral choice task was defined as selecting the four stimuli directly or indirectly related to positive images during training (i.e., S1, O1, O2 and O5). Choosing any other stimulus (S2, O3, O4 or O6) was defined as an incorrect response. Twenty participants (63%) who passed the equivalence test selected all of the correct stimuli while a further twelve (37%) failed to do so. In contrast, only a single participant who failed the equivalence test (6%) selected the correct stimuli while fourteen (88%) selected the incorrect options (the computer failed to record the behavioral choice data for a single participant). When a Fisher's exact test was performed to examine whether stimulus selection was distributed differently across those who passed versus failed the equivalence test a significant effect was obtained ($p < .001$). This indicates, based on the odds ratio, that the likelihood of selecting S1, O1, O2 and O5 was 23 times higher if participants passed the equivalence test than if they failed to do so.

Discussion

Experiment 4 examined a second potential boundary condition of our account – namely - would changes in liking due to intersecting regularities still be observed as the complexity of those intersections increase? Stated another way, would a change in liking still be observed when those changes depend on both *direct* and *indirect* intersections between contingencies? To test this assumption we established a number of operant contingencies that directly intersected in terms of a common antecedent stimulus (i.e., *S1-R1-O1* and *S1-R2-O2* vs. *S2-R3-O3* and *S2-R4-O4*). We then established an addition set of contingencies wherein the outcome of one intersected with the antecedent stimulus of another (*positive stimulus (S3)-R5-O5* and *O5-R5-O2* vs. *negative stimulus (S4)-R7-O6* and *O6-R8-O4*). In this way, we established a series of contingencies that directly and indirectly intersected with one another. Results revealed that participants not only acted as if the stimuli in those intersecting contingencies were equivalent to one another (i.e. *Positive stimuli (S3)-O5-O2-O1-S1* vs. *Negative stimuli (S4)-O6-O4-O3-S2*) but also showed evidence of a transfer of valence from the positive and negative stimuli to the other stimuli in those respective relations. That is, they responded positively to O5, O2, O1 and S1 and responded negatively to O6, O4, O3 and O2. Corresponding evaluative effects were also evident on automatic (traditional IAT) and behavioral approach tasks. Similar to Experiments 2 and 3, this pattern of evaluative responding was demonstrated by participants who passed the equivalence test and was largely absent in those who failed.

General Discussion

Given the central role that evaluation plays in phenomena such as decision making, attention, and memory, researchers have looked for ways to predict and influence the former in order to better understand the latter. Until now, researchers have often sought to change likes and dislikes via regularities in the presence of a single stimulus (mere exposure), two or more stimuli (EC), or behavior and its consequences (approach/avoid learning). The current research set out to determine whether intersections between regularities represent a previously unrecognized class of procedures for changing liking. Our results provide strong support for this hypothesis and suggest that when environmental regularities such as operant contingencies intersect with one another, the evaluative properties of stimuli involved in those intersections can change. These changes in liking were observed across a range of stimuli and procedures and were evident when self-report measures, implicit measures (IAT) and behavioral choice measures were employed. In what follows, we consider several important properties of evaluative responses that emerge in this way. We will also analyze these effects in functional and mental terms. Finally, a number of open questions and future directions are highlighted.

Intersecting Regularities: Two Core Properties

To the best of our knowledge, the above research represents the first attempt to systematically explore whether intersecting regularities can lead to changes in liking. Given the staggering number of ways in which regularities can potentially intersect, we decided to restrict our initial efforts to a sub-set of all possible intersections between just one type of regularity (i.e., operant contingencies). The results from Experiments 1-4 indicate that changes in liking due to such intersections are characterized by two key properties.

Generativity. First, many evaluative responses that resulted from intersecting regularities emerged in the absence of direct pairings or instruction. Take Experiment 1. Participants learned that emitting a specific response in the presence of an antecedent led to the presentation of either a neutral (O1) or valenced (O2) outcome (i.e. *S1-R1-O1*and *S1-R1-O2*). Thereafter participants evaluated outcomes that intersected with contingencies containing positive stimuli positively and outcomes that intersected with contingencies containing negative stimuli negatively. This is despite the fact that valenced and neutral stimuli were never paired at any point in time. Likewise, participants in Experiments 2-3 learned that emitting a response in the presence of a valenced stimulus (S1) produced one outcome (O1) while emitting another response in the presence of a neutral stimulus (S2) led to that same outcome (*S1-R1-O1* and *S2-R2-O1*). This intersection in terms of a common outcome led to a change in liking towards stimulus (S2) despite the fact that it was never directly related with valenced images at any point in time. Therefore, when stimuli participate in intersecting regularities, people may act as if those stimuli are mutually substitutable or equivalent in some way even though they were never paired with, or directly related to, one another in the past.

Versatility. A second characteristic of intersecting regularities is that it represents a highly versatile means of engineering likes and dislikes. In the above studies, stimuli consistently acquired evaluative properties regardless of the number, nature and complexity of the intersections involved. Changes in liking were observed when contingencies intersected in terms of just one or two elements and these intersections could involve any potential type or combination of elements. For instance, learning that two operant contingencies shared a common outcome (Experiments 2-3), stimulus (Experiment 4), or stimulus and response (Experiment 1) led to a transfer of evaluative properties from one (valenced) stimulus to another (neutral). Increasing the complexity of the intersection, so that changes in liking required direct and indirect intersections led to similar outcomes as seen at lower levels of complexity (Experiment 4). Evaluative learning effects were still evident even when attention was directed away from evaluation and towards an irrelevant domain (Experiment 3), and we observed no evidence that our effects were influenced by demand compliance or conversational norms. Thus the above results provide converging evidence that different numbers, types and combinations of intersections can all give rise to novel evaluative responses.

Effects of Intersecting Regularities: A Theoretical Analysis

As we outlined in the introduction, stating that an observed change in liking is the product of intersecting regularities merely implies that the change in liking is a product of specific factors in the environment of the individual. When approached from this perspective we see that there are two different levels at which to analyze the effects of intersecting regularities: (1) a functional level that aims to describing the elements in the environment that *moderate* these effects; and (2) a mental level that aims to uncover the mental mechanisms that *mediate* the impact of intersecting regularities on liking. In the following section, we turn our attention to possible functional and mental analyses of this phenomenon.

The functional-analytic level of analysis. One of the core findings in the current paper is that self-reported and automatic evaluative responding emerged via intersecting regularities only when participants treated stimuli as being equivalent in other ways (i.e., when they matched them in novel and untrained ways during an MTS task). At the functional-analytic level, this would seem to suggest that the changes in evaluative responding observed in Experiments 1-4 represent an interaction between proximal regularities in the current environment (i.e., intersecting regularities) and distal regularities in the organisms past environment (i.e., those that allow people to act as if stimuli are equivalent to one another in some way). For instance, equivalence responding (like other types of relational responding) refers to a type of generalized operant behavior that seems to emerge during early childhood. It is established via an extensive history of learning for responding to the functional relationship between stimuli regardless of their physical properties. For instance, if a child learns that the written word ‘poison’ (A) is the same as a picture containing an unknown symbol (B) and that the latter is the same as the spoken word ‘G-I-F’ (C), they will likely avoid consuming any items that contain images of B or that are labeled “GIF”. In other words, they will spontaneously act as if the above stimuli are related in four new ways based on two directly trained relations (i.e., that C is the same as A; A the same as C; B the same as A and C the same as B). This occurs even though the word, picture and sound bear no physical similarity to one another in addition to the fact that the child has never directly experienced the consequences of consuming the latter two stimuli in the past (for book length treatments of equivalence and other types of relational responding see Dymond & Roche, 2013; Hayes et al., 2001; see also De Houwer & Hughes, 2015; Hughes, De Houwer, & Barnes-Holmes, 2015, for similar arguments in the context of evaluative conditioning).

The above example serves to highlight an important point: changes in the psychological properties of a stimulus (in this case liking) can be just one part of the wider class of behaviors that involve “acting as if” stimuli are equivalent to one another. If changes in liking due to intersecting regularities are instances of equivalence responding, than these changes should coincide with other changes that are also part of equivalence responding, such as relating stimuli in entirely novel and untrained ways that are in-line with what they have previously learned. We observed evidence for this in each and every experiment where a test for equivalence was included (Experiments 2-4). At the same time, if intersecting regularities give rise to equivalence responding, they should give rise not only to a transfer of evaluative stimulus properties (i.e., liking) but also a transfer of non-evaluative stimulus properties. Ongoing work in our labs corroborate this idea as well. We have found that discriminative functions can also transfer between stimuli that participate in intersecting regularities. More specifically, when people learned that an antecedent stimulus from one regularity signals that a certain response should be emitted they also act as if stimuli from a second intersecting regularity also signals that same response. The fact that different kinds of functions can transfer between stimuli that are part of intersecting regularities fits with a large body of research in the human learning literature, wherein a spectrum of psychological properties (including evaluation and discrimination) have been found to transfer from one stimulus to another in an equivalence relation (e.g., Dougher, Hamilton, Fink, & Harrington, 2007; Roche, Kanter, Brown, Dymond, & Fogarty, 2008; Valdivia-Salas, Dougher, & Luciano, 2013).

Our results also have wider implications for functional psychology. From a functional perspective, our findings demonstrate that equivalence relating can be brought about in a previously unrecognized manner. In most studies in this area, equivalence relations are established via Matching-To-Sample tasks (i.e., selecting one of several target stimulus on the basis of a sample stimulus) or on the basis of stimulus pairings (Leader & Barnes-Holmes, 2001; although see Dymond & Whelan, 2010). Our results suggest that intersections between operant contingencies can also bring about such a pattern of responding, in which changes in liking coincide with other aspects of stimulus equivalence. Indeed, like the MTS task, the procedure we used to establish intersecting operant contingencies seems to represent a context in which participants are required to make conditional discriminations between different responses (in the presence of certain antecedents) and these responses lead to certain outcomes[11]. In Experiment 1, for example, participants had to conditionally discriminate between two responses (R1 and R2) in the presence of one of two antecedents (S1 and S2) and these responses led to different outcomes (i.e., the presentation of positive stimuli or O1 versus negative stimuli and O2). Likewise, in Experiments 2-3, participants discriminated between a number of different responses in the presence of valenced or non-valenced antecedents and these responses led to the presentation of a single stimulus. In other words, it may be that procedures designed to establish intersections between regularities represent a context signaling that stimuli should be related as similar or equivalent to one another (see Zentall, Wasserman, & Urcuioli, 2014, for other procedures that establish intersecting regularities). Once stimuli are related in this way the evaluative properties of the stimuli and responses involved may change in-line with the relation established. This would explain why performance on the equivalence test was linked to the presence or absence of evaluative responses in each of our studies[12].

Alternative functional explanations. Taking a step back, it may be tempting to describe the changes in liking observed in Experiments 1-4 as instances of other, more basic

[11] Note that the overlap between the current design and MTS tasks only applies for intersections that take place between operant contingencies and not to intersections between other types of regularities.
[12] A reviewer noted that the observed relation between learning via intersecting regularities and equivalence test performance might have reflected inter-individual differences in general learning. For instance, it might be that certain individuals are better learners than others insofar as they pay closer attention to stimuli during the learning phase. If so, then they would be expected to show larger effects on all indices of learning, whether it is the contingency memory test, equivalence test or evaluative measures. While certainly possible, it is unlikely that this “better learner” explanation is valid for our results given that participants progressed from the learning phase to the evaluative measures in Experiments 3-4 after having emitted either (a) 30 out of 36 non-consecutively correct, or (b) 16 consecutively correct responses during the final block of training trials.

behavioral principles such as classical conditioning. Consider the results of Experiment 1 (see Figure 2). Perhaps the valence of the positive outcome was transferred to the common response (i.e., backward conditioning) which in turn gave rise to a change in the valence of a neutral stimulus (i.e., forward conditioning). Now consider Experiments 2 and 3 (see Figure 3). It may be that the valence of the positive stimulus was transferred to the first response, which was in turn transferred to the outcome (i.e., forward conditioning). Thereafter, valence might have been transferred to the second response and then to the neutral stimulus (i.e., two successive instances of backward conditioning). An even more complex description in terms of forward and backward conditioning could – in principle – be provided for the findings of Experiment 4. Yet these accounts in terms of chains of individual regularities seem to encounter a number of immediate difficulties. First, one has to assume that valence transfers in different steps as the result of different regularities in the presence of two events. Although such a “chain of changes” in valence (i.e., A changes B because of A-B pairings, B changed C because of B-C pairings) might actually occur, there are reasons to believe that our effects are not instances of such a chain of evaluative conditioning effects. First, we are not aware of any previous studies showing changes in liking when chains of pairings involve more than two steps. We do know, however, that conditioning in general becomes weaker when more steps are involved in a chain of pairings. Nevertheless, in our studies, substantial changes in liking occurred even when multiple steps of pairings were involved (e.g., Experiment 4). Second, the available evidence suggests that, like other types of conditioning effects, evaluative conditioning is generally weaker when the valenced stimulus (US) always occurs before the originally neutral stimulus (CS; i.e., backward evaluative conditioning) relative to when the valenced stimulus is presented after the originally neutral stimulus (see Hofmann et al., 2010, for a review). In our studies, changes in liking occurred even though they involved (multiple) backward relations. Third, to explain our results in terms of a chain of evaluative conditioning effects, one has to assume that valence can transfer from stimuli to responses and back, irrespective of the order of these events. At least currently, there is little or no evidence for such a phenomenon. Finally, a description of our effects in terms of simple evaluative conditioning cannot encompass the fact that the observed changes in liking only occurred when participants passed a simple test for equivalence responding[13].

Explanations at the mental level of analysis. Cognitive and social psychologists typically explain evaluative learning effects in one of three ways: in terms of the formation, storage and activation of (a) associative links between mental representations in memory, (b) propositions that specify the content and structure of relations between stimuli and events or (c) some combination of the two (dual-process models). Importantly, these different types of models vary in their capacity to explain the current pattern of results. We therefore discuss the implications of these models separately in the next three sections.

Associative mental models. Associative learning models that argue for unidirectional association formation fail to offer a satisfactory explanation of our main findings. Consider Experiment 1 wherein the same stimulus and response gave rise to two difference outcomes (either a valenced or neutral outcome). From an associative perspective, a number of S-R and R-O associations may have emerged during the training phase such that the same response (R1) caused people to think of two different outcomes (O1 and O2). Given a sufficient number of training trials a mental association may have been formed between O1 and O2 and it was this association that mediated the observed change in liking. However, unidirectional

[13] One could argue that sensory preconditioning (i.e., a change in liking of CS1 following CS1-CS2 pairings followed by CS2-US pairings) is an instance of evaluative learning via intersecting regularities. Note that such conceptualization of sensory preconditioning is indeed more plausible than a conceptualization in terms of evaluative transfer via chains of CS-US pairings. A chain interpretation is problematic given that CS2 is neutral at the time of CS1-CS2 pairings.

association models cannot explain the outcomes obtained in Experiments 2-4. Take Experiments 2 and 3 wherein two operant contingencies intersected in terms of a common outcome (*valenced stimulus (S1)-R1-O1* and *neutral stimulus (S2)-R2-O1*). If association formation only occurs in a forward fashion then participants should not have come to like or dislike S2. Yet this was clearly not the case (see also Experiment 4).

Associative learning models which do allow for bidirectional mental associations between antecedents, responses and outcomes could successfully predict all of the observed changes in liking (see Custers & Aarts, 2011; de Wit & Dickinson, 2009; Elsner & Hommel, 2004). Nevertheless, this class of models fails to explain why changes in liking only emerged when participants passed a test for equivalence. Individuals who failed to act as if stimuli were equivalent to one another showed no evidence of self-reported or automatic evaluative responding despite the fact that (a) they were exposed to a similar number of training trials as those who passed the equivalence test and (b) both the pass and fail groups exited the learning phase with near perfect response accuracy (also see Footnote 8). Thus it appears that something more than simple associations is needed in order to explain the current outcomes – otherwise those who passed or failed the equivalence test should have produced similar effects. In other words, unidirectional and bidirectional associative models appear to be problematic on different grounds.

Propositional models. Whereas the current data undermine associative accounts, they provide firm support for propositional models that involve qualified links between mental representations in memory. Whereas associations simply convey the strength with which representations are linked, propositions specify their strength, structure and content (e.g., '*X is opposite to Y*' or '*X is larger than Y*'). Likewise, while associations gradually develop with many experienced pairings, propositions can be formed on the basis of experience, instructions, inference or deductive reasoning (De Houwer, 2009, 2014). When combined these two properties of propositions may account for the current set of findings.

To illustrate, consider Experiment 1. During the training phase participants may have formed propositions concerning the relationship between stimuli and events based on their experience (e.g., "*If I press R1 when I see S1 then either positive images (O1) or neutral images (O2) will appear"*). This proposition may have served as the foundation upon which novel propositions were subsequently inferred (e.g., "*O1 is the same as O2*"), and it was these inferred propositions which mediated the obtained effect. A similar explanation may also account for Experiments 2-4. In the latter case, participants may have generated a set of propositions based on experience (e.g., "*If I press R1 when I see positive stimuli (S1) or R2 when I see neutral brands (S2) then a Chinese ideograph (O1) will appear*'). Once again, these propositions may have then given rise to an additional set of inferred propositions (e.g., '*positive images (S1) are the same as neutral brand names (O1 and neutral brand names are the same as Chinese ideographs (O1))*). It may be that these inferred propositions – rather than those that arise from experience – mediated the observed changes in liking (e.g., "*the brand name (S2) and ideograph (O1) are Good"*).

The fact that changes in liking emerged only when participants passed the equivalence test is consistent with the above explanation. The equivalence test was designed to test whether people will 'act as if' stimuli are related to one another in untrained ways. It seems likely that performance on this task was mediated by the inferred propositions outlined above. Specifically, performance during the equivalence test seems to require participants infer a series of novel propositions from a limited set of propositions that arise on the basis of direct experience. Inferential reasoning provides a mechanism by which these novel propositions might be generated and via which these novel behaviors can thus arise.

Our findings also provide support for the notion that, once formed, propositions may be activated automatically from memory and mediate implicit measures of liking (see De Houwer, 2014; Hughes, Barnes-Holmes & De Houwer, 2011). Automatic evaluative responses in Experiments 2-4 emerged for stimuli that were never directly paired with valenced images but only when participants passed the equivalence test. This can be explained by assuming that propositions established during training were stored in memory and automatically retrieved during the IAT.

Dual process models. Finally, dual-process accounts that allow for rules (Smith & DeCoster, 2000), judgments (Kahneman, 2003), and propositions (Gawronski & Bodenhausen, 2011) to feed into and create novel associations could also account for our findings – provided that certain pre-conditions are met. For instance, and similar to the above propositional model, direct experience with a number of operant contingencies may have led to the formation of propositions about those experiences, and in turn, to the generation of additional, inferred propositions about the original set of propositions. If correct, then these inferred propositions could have been transformed into and stored as associations in memory, explaining the observed pattern of self-reported and automatic evaluations. To some extent, it is unsurprising that, as a category, dual-process models can account for our results at least as good as single-process propositional models. For any version of a single-process propositional model, one could simple add an associative system to create a dual-process model that explains at least as much as that single-process propositional model. Hence, the real value of dual-process models needs to be evaluated on the basis of what it adds to models that include only one of the processes.

Open Questions and Future Directions

Our studies open up an entirely new line of research on evaluation that we have only begun to explore. We did not consider every possible way in which operant contingencies can intersect nor did we examine other intersections between regularities that involve non-operant contingencies. Future work could determine whether changes in liking also emerge when different types, numbers and combinations of regularities intersect with one another. It may be that repeated presentations of a single stimulus (mere exposure) leads to a change in liking towards a neutral stimulus when they are paired together (EC). Likewise, a previously neutral stimulus may reinforce a particular pattern of responding in an operant contingency after (a) being repeatedly presented on its own (mere exposure) or (b) previously paired with an appetitive stimulus (EC).

Critically, our findings suggest that the way in which intersections are established will influence the probability of observing changes in liking. Take the learning phases of Experiments 1-4 wherein participants were exposed to S-R and R-O relations both of which were part of the intersecting regularity. In order to complete the learning phase only the S-R relations were relevant. That is, on each trial participants could select the correct response merely on the basis of the stimulus. This aspect of the procedure might have resulted in decreased attention for the R-O relations and thus the magnitude of the learning effect. This idea fits with the fact that participants who failed the contingency memory test (which probed for R-O relations) also showed little evidence for evaluative responding. It therefore seems likely that certain environmental moderators (e.g., task instructions) and mental mediators (e.g., attention; epistemic knowledge) will influence the probability of observing relational learning via intersecting regularities[14].

At the same time, we currently know very little about the functional properties of these effects, from their resistance to extinction to their susceptibility to reversal via counter-conditioning or re-evaluation. Our knowledge of which aspects of the intersection (e.g., number of training trials), stimulus (e.g., modality or valence identity), measurement procedure (explicit or implicit), organism (human versus non-human), context (conditions that promote awareness or unawareness) or induction method (direct experience, instructions, inferences or observation) that are necessary and sufficient to produce these effects is limited. For instance, it may be that a change in liking occurs only when participants are aware of the intersection between regularities and that extensive training is necessary to establish such awareness. It may also be the case that instructions which specify an intersection give rise to similar outcomes compared to when people directly experience events that incorporate those intersections for themselves.

We also know that humans are capable of relating stimuli and events in a diverse number of ways, from equivalence (‘*X is the same as Y*’) to opposition (‘*X is the opposite of Y*’), temporally (‘*X comes before Y*’) and causally (‘*X causes/prevents Y*’) not to mention hierarchically (‘*X is part of Y*’) and deictically (‘*X is related to me*’) (for a detailed overview

[14] Participants in Experiments 1-4 were always required to respond to stimuli that were part of intersecting operant contingencies using the same hand. It may thus be that the hand with which they responded was an additional intersection between regularities that contributed to the observed changes in liking. Although this may be true for the current work and although this would not undermine the general conclusion that liking can change as the result of intersecting regularities, other studies show that intersections with regard to the hand used are not crucial in order to obtain changes in liking. In the self-referencing task, for example, the hand with which one responds to stimuli alternates across blocks. During some blocks participants respond to ‘self’ and neutral word 1 using their right hand while responding to ‘others’ and neutral word 2 using their left hand. These response requirements are then reversed in the next block so that people respond to self and neutral word 1 using their left hand and others and neutral word 2 with their right hand. Effects obtained from this task are not influenced by this method factor.

see Hughes & Barnes-Holmes, in press-a). One possibility is that when people encounter intersections between regularities their first response is to treat stimuli in those regularities as equivalent to one another. Indeed, this is precisely what we observed in Experiments 1-4. Yet there is no *a priori* reason why elements of intersecting regularities cannot be related in non-equivalent ways as well. Stated another way, when elements of intersecting regularities signal that other elements are ‘opposite’, ‘more/less than’, one another, or that they ‘cause/prevent’, ‘come before/after’ or ‘are part of ’ one another, then entirely different (but predictable) patterns of valence change should be observed. To illustrate this more clearly, imagine that you perform a training phase that consists of a series of blocks. Whenever the computer screen is blue, two randomly selected positive stimuli are assigned to one key whereas two randomly selected negative stimuli are assigned to a second key. In blocks where the screen is green, however, the two stimuli assigned to the same response have an opposite valence. As a result of this pre-training, the background color of the computer screen becomes a contextual cue that signals how stimuli that are part of intersecting regularities (i.e., that are assigned to same key) are related in terms of valence (i.e., blue signals equivalence whereas green signals opposition). When during a subsequent learning phase, a neutral stimulus is assigned to the same key as a positive stimulus, the neutral stimulus might become positive if the background during the learning phase is blue whereas the same events might result in a negative valence of the originally neutral stimulus when the background color during learning is green (see Whelan & Barnes-Holmes, 2004 for a similar impact of contextual cues on relational responding in matching-to-sample tasks). Future research could unpack this issue and explore whether intersecting regularities also allow for stimuli to be related in non-equivalent ways and whether these relations also lead to stimuli being evaluated in non-equivalent ways.

Finally, while the current work speaks only to the formation of novel likes and dislikes it is entirely possible that intersecting regularities can also be used to modify or even eliminate pre-existing preferences as well. To illustrate, imagine that responding in the presence of a single antecedent stimulus leads to either an image of an outgroup member or a positive stimulus. It may be that following this training the evaluation of outgroup members becomes less negative or even positive. Similar outcomes have already been obtained in the domain of EC (e.g., counterconditioning; Olson & Fazio, 2006), approach/avoidance training (Kawakami et al., 2007) as well as in the context of the self-referencing task (Prazienkova, Paladino, Zogmaister, Perugini, & Richetin, 2014). Evidence that intersecting regularities are capable of establishing and eliminating preferences would provide further support that this way of arranging the environment provides a viable means of generating and manipulating likes and dislikes.

Although the primary focus of this paper concerns evaluative learning, intersecting regularities may lead to a change in a diverse spectrum of non-evaluative properties above and beyond those observed here. Above, we noted that we already observed a transfer of discriminative properties as the result of intersecting regularities. It may be that intersections can also produce other changes, such as changes in accessibility, attention, memory, motivation, emotion and interpretation. If correct, then intersecting regularities may provide an experimental means of altering attention, memory, interpretation and judgment in the absence of direct experience (see Richetin, Perugini, & Mattavelli, 2015 for preliminary work in this vein). Moreover, and in addition to potentially novel effects, it may be that previously identified phenomena (typically labeled using different terms) also represent instances of learning via intersecting regularities. For example, task rule congruency effects in the cognitive control literature involve verbally instructed or directly experienced operant contingencies that intersect with one another in terms of a common stimulus, responses or outcome (Liefooghe, Wenke, & De Houwer, 2012). When these intersecting contingencies involve common responses or outcomes “congruency effects” are typically observed; yet when those same contingencies involve different response and outcomes “incongruency effects” result. Goal-directed learning research also shows that people are often slower to learn when certain types of operant contingencies intersect compared to when they do not (e.g., De Wit, Niry, Wariyar, Aitken, & Dickinson, 2007). There may be many other effects and phenomena that can be conceptualized in this way that we have yet to identify. The point here is that conceptualizing the above effects as instance of intersecting regularities has several advantages: it equips researchers with a means of identify similarities and differences across seemingly unrelated phenomena (i.e., it has heuristic value) which may in turn lead to new insights in each of these domains (i.e., it has predictive value).

Conclusion

The work reported in this paper adds to our knowledge about how likes and dislikes are acquired and how they can be changed. Its main contribution resides in the proposal of intersecting regularities as a new class of procedures for changing liking and thereby in establishing evaluative learning via intersecting regularities as a new type of evaluative learning effect. We do not argue that we are the first to implement intersecting regularities in a procedure or to examine the impact of an intersecting regularity on behavior. In fact, the starting point of our endeavor was an existing demonstration of what could be considered as the effect of a particular instance of intersecting regularities on liking (i.e., the self-referencing effect). We are, however, the first to recognize intersecting regularities as a separate class of procedures that could give rise to changes in liking. Without this essential ingredient, we would not have been able to conceive of the possibility that all kinds of previously untested instances of intersecting regularities might also produce changes in liking.

Our findings suggest that intersections between regularities can be used to engineer self-reported and automatic evaluative responses and that these emerge only when people treat the stimuli involved in those intersections as equivalent to one another. This work offers a number of general predictions about changes in liking that occur when regularities intersect and also sets the stage for a host of new questions about how factors such as the regularity, intersection, stimulus, procedure, organism, context, and induction method moderate these outcomes (i.e., it has predictive value). Our work also has heuristic value insofar as it highlights a new type of evaluative learning and provides a better means of organizing existing effects. Thus the concept of intersection regularities may strengthen the study of evaluative learning by organizing existing and generating new knowledge about the conditions under which likes and dislikes emerge, persist and change.

References

Augustson, E.M. & Dougher, M.J. (1997). The transfer of avoidance evoking functions through stimulus equivalence classes. *Journal of Behaviour Therapy and Experimental Psychiatry, 28,* 181 -191.

Bohner, G., & Dickel, N. (2011). Attitude and attitude change. *Annual Review of Psychology, 62,* 391-417.

Custers, R., & Aarts, H. (2011). Learning of predictive relations between events depends on attention, not on awareness. *Consciousness and Cognition, 20,* 368-378.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230–241.

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6,* 202 -209.

De Houwer, J. (2014). A propositional perspective on context effects in human associative learning. *Behavioural processes, 104,* 20-25.

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin & Review, 20*, 631 -642.

De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology, 43,* 972–978.

De Houwer, J., & Hughes, S. (2015). Why is evaluative conditioning important? On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. Social Cognition. Manuscript submitted for publication.

de Wit, S., & Dickinson, A. (2009). Associative theories of goal-directed behaviour: A case for animal-human translational models. *Psychological Research, 73,* 463–476.

de Wit, S., Niry, D., Wariyar, R., Aitken, M. R. F., & Dickinson, A. (2007). Stimulus -outcome interactions during conditional discrimination learning by rats and humans. *Journal of Experimental Psychology: Animal Behavior Processes, 33*(1), 1 –11.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86,* 345-355.

Dougher, M. J., Hamilton, D., Fink, B., & Harrington, J. (2007). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the Experimental Analysis of Behavior, 88(2),* 179-197.

Dymond, S. & Rehfeldt, R. (2000). Understanding complex behaviour: The transformation of stimulus functions. *The Behavior Analyst*, *23*, 239-254.

Dymond, S., & Roche, B. (Eds.), (2013). *Advances in Relational Frame Theory: Research and Application*. Oakland, CA: New Harbinger.

Dymond, S., & Whelan, R. (2010). Derived relational responding: A comparison of matching to sample and the relational completion procedure. *Journal of the Experimental Analysis of Behavior, 94,* 37-*55*.

Ebert, I. D., Steffens, M. C., von Stülpnagel, R., & Jelenec, P. (2009). How to like yourself better, or chocolate less: Changing implicit attitudes with one IAT task. *Journal of Experimental Social Psychology, 45,* 1098–1104.

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action effect learning. *Psychological Research, 68,* 138-154. DOI: 10.1007/s00426-003-0151-8.

Gast, A., De Houwer, J., & De Schryver, M. (2012). Evaluative conditioning can be modulated by memory of the CS–US pairings at the time of testing. *Learning and Motivation, 43(3),* 116-126.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44,* 59–127.

Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology, 39(1)*, 16-30.

Galdi, S., Arcuri, L., & Gawronski, B. (2008). Automatic mental associations predict future choices of undecided decision-makers. *Science, 321*, 1100-1102.

Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the Implicit Association Test. *Journal of Consumer Research, 35,* 178 –188.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197-216.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136,* 390–421.

Hollands, G. J., Prestwich, A. & Marteau, T. M. (2011). Using aversive images to enhance healthy food choices and implicit attitudes: An experimental test of evaluative conditioning. *Health Psychology, 30(2),*195-203.

Hughes. S., De Houwer, J. & Barnes-Holmes, D. (2015). The Moderating Impact of Distal Regularities on the Effect of Stimulus Pairings: A Novel Perspective on Evaluative Conditioning. Manuscript submitted for publication.

Hughes, S., & Barnes-Holmes, D. (in press-a). Relational Frame Theory: The Basic Account. In S. Hayes, D. Barnes-Holmes, R. Zettle, and T. Biglan (Eds.), Handbook of Contextual Behavioral Science. New York: Wiley-Blackwell.

Hughes, S., & Barnes-Holmes, D. (in press-b). Relational Frame Theory: Implications for the Study of Human Language and Cognition. In S. Hayes, D. Barnes-Holmes, R.

Zettle, and T. Biglan (Eds.), Handbook of Contextual Behavioral Science. New York: Wiley-Blackwell.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The Dominance of Associative Theorising in Implicit Attitude Research: Propositional and Behavioral Alternatives. *The Psychological Record*, *61*, 465–498.

Kahneman, D. (2003). A perspective on judgement and choice. *American Psychologist, 58*, 697–720.

Kawakami, K., Phills, C. E., Steele, J. R. & Dovidio, J. F. (2007). (Close) Distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology, 92,* 957 -971.

Leader, G., & Barnes-Holmes, D. (2001). Matching-to-sample and respondent-type training as methods for producing equivalence relations: Isolating the critical variable. *The Psychological Record, 51*, 429-444.

Liefooghe, B., Wenke, D., & De Houwer, J. (2012). Instruction-based task-rule congruency effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*, 1325-1335.

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behavior Research and Therapy, 1*, 57-102.

Moreland, R. L., & Topolinski, S. (2010). The Mere Exposure Phenomenon: A Lingering Melody by Robert Zajonc. *Emotion Review, 2(4)*, 329-339.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, 86,653-667.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically-activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32,* 421-433 .

Payne, B. K., Cheng, S. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277–293.

Perugini, M., Zogmaister, C., Richetin, J., Prestwich, A., & Hurling, R. (2013). Changing implicit attitudes by contrasting the self with others. *Social Cognition , 31,* 443-464.

Prazienkova, M., Paladino, M.P., Zogmaister, C., Perugini, M., & Richetin, J. (2014). *Reducing out-group infra-humanization and implicit intergroup bias through self-referencing*. Working manuscript.

Prestwich, A., Perugini, M., Hurling, R., & Richetin, J. (2010). Using the self to change implicit attitudes. *European Journal of Social Psychology, 40,* 61–71.

Roche, B. T., Kanter, J. W., Brown, K. R., Dymond, S., & Fogarty, C. C. (2008). A comparison of "direct" versus "derived" extinction of avoidance responding. *The Psychological Record, 58,* 443-464.

Röhner, J., Schröder-Abé, M., & Schütz, A. (2013). What do fakers actually do to fake the IAT? An investigation of faking strategies under different faking conditions. *Journal of Research in Personality, 47,* 330-338.

Schwartz, N. (1994). Judgement in a social context: Biases, shortcomings, and the logic of conversation. *Advances in experimental social psychology, 26*, 123-162.

Sidman, M. (2009). Equivalence relations and behavior: An introductory tutorial. *The Analysis of Verbal Behavior, 25,* 5–17.

Smith, C., De Houwer, D., & Nosek, B. (2013). Consider the source of persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin, 39(2)*, 193-205.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4,* 108–131.

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology*, *71*, 888-898.

Valdivia-Salas, S., Dougher, M., & Luciano, C. (2013). Derived relations and generalized alteration of preferences. *Learning and Motivation, 41(2)*, 205-217.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition and Emotion, 19,* 175-196.

Whelan, R., & Barnes-Holmes, D. (2004). The transformation of consequential functions in accordance with the relational frames of same and opposite. *Journal of Experimental Analysis of Behavior*, *82(2),* 177-195.

Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual review of psychology, 51(1),* 539-570.

Zentall, T. R., Wasserman, E. A., & Urcuioli, P. J. (2014). Associative concept learning in animals. *Journal of the Experimental Analysis of Behavior, 101*, 130-151.